Singular geodesic coordinates for representing diffeomorphic maps in computational anatomy, with application to the morphometry of early Alzheimer's disease in the medial temporal lobe

by

Daniel Jacob Tward

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March 17, 2017

Abstract

In this work we develop novel algorithms for building one to one correspondences between anatomical forms by providing a sparse representation of dense registration information. These sparse parameterizations of complex high dimensional data allow robustness in the face of noise and anomalies, and a platform for inference that is effective in the face of multiple comparisons. We review background in the theory of generating smooth, invertible transformations (the diffeomorphism group), and build our parameterization as a function supported on surfaces bounding anatomical structures of interest. We show how dimensionality can be reduced even further and still provide a rich family of mappings using principal component analysis or Laplace Beltrami eigenfunctions supported on the surface. We develop algorithms for surface matching and image matching within this model, and demonstrate the desired robustness by working with published large neuroimaging datasets that include many low quality examples. Finally we turn to addressing challenges associated with some specific data types: images with multiple labels, and longitudinal data. We use the mapping tools developed to draw conclusions about the progression of early Alzheimer's disease in the medial temporal lobe.

Acknowledgments

I want to thank my mentor, Dr. Michael Miller, for providing me with an environment where I could explore my interests and be creative, and for giving me the guidance to make a contribution to neuroscience and computational anatomy.

I'd also like to thank my thesis committee members. Dr. Laurent Younes for invaluable discussions about mathematical details, as well as for teaching courses that piqued my interest in variational methods and statistics. Dr. Susumu Mori for providing the resources for obtaining beautiful and interesting neuroimaging data.

This work couldn't have been completed without help from my colleagues at the Center for Imaging Science. Chelsea Sicat and Timothy Brown for their tireless work performing manual segmentations. Dr. Tilak Ratnanather for inspiration. And Anthony Kolasny for keeping computational resources working.

Last I'd like to thank my friends and family, classmates and fellow musicians, for contributing so much to the well being necessary for a sustained effort over the years.

Daniel Jacob Tward
Baltimore, Maryland
March 17, 2017

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurocognitive disorder, usually occurring in middle or late life. McKhann et al. [1] formalized diagnostic criteria for Alzheimer disease in 1984, stating that *definite* AD can be confirmed only by histopathologic evidence (presence of "plaques and tangles") at biopsy or usually autopsy. The clinical criteria for *probable* AD includes dementia (established for example through the Mini-Mental Test [2]), cognitive deficit, and worsening memory.

In 1995 Braak and Braak [3] described deposition of extracellular amyloid, and intracellular neurofibrillary changes (plaques, tangles, and threads). They found tangles and threads have low variability between patients, and could be used for staging disease progression. They defined 6 stages

1-2. Transentorhinal stage: Cell layers 1 and 2 are mainly affected.

3-4. Limbic stage: The entorhinal and transentorhinal cortex, CA1 of the hippocampus, and some isocortex is affected.

5-6. Isocortical stage: Virtually all association areas are affected.

Since they worked with autopsied brains they did not attempt to correlate their findings with clinical symptoms.

In 2011, a modern definition of AD was established which includes a discussion of biomarkers and disease staging. Probable AD was divided into three clinical stages.

1. Preclinical AD was established by Sperling et al. [4]. In this stage measurable biomarker changes are present before clinical symptoms occur.
2. Mild cognitive impairment was described by Albert et al. [5]. Here there are measurable changes in mental status tests, but not disruptive to life.
3. Finally, dementia due to AD is described by McKhann et al. [6].

In summarizing these diagnostic criteria, Jack et al. [7] divides biomarkers into two categories

1. The biomarkers of Aβ accumulation, which generally present first .
2. The biomarkers of neuronal degeneration or injury, which likely follow causally from the previous category

Despite careful characterization, these biomarkers are largely ignored for clinical purposes. At the preclinical stage Sperling et al. [4] states:

> This framework is not intended to serve as diagnostic criteria for clinical purposes. Use of these biomarkers in the clinical setting is currently unwarranted because many individuals who satisfy the proposed research criteria may not develop the clinical features of AD in their lifetime

At the MCI stage Albert et al. [5] states:

> More research needs to be done to ensure biomarkers are appropriately designed, standardized from one locale to another including cut-points for diagnosis, access to biomarkers is limited

At this stage biomarkers can establish support for the underlying etiology of the clinical syndrome, and determine the likelihood of progression. In particular they suggest studying hippocampal volume or medial temporal atrophy, by volumetric measures or visual rating.

At the AD dementia stage McKhann et al. [6] notices:

> Disproportionate atrophy on structural magnetic resonance imaging in medial, basal, and lateral temporal lobe, and medial parietal cortex

and yet states:

> We do not advocate the use of AD biomarker tests for routine diagnostic purposes.

Work continues in characterizing biomarkers, and Duara et al. [8] provides a recent review of the importance of positron emission tomography (PET) and structural MRI to quantify them. These include amyloid PET imaging, decreased fluorodeoxyglucose (FDG) uptake in PET in pattern involving temporoparietal cortex, atrophy on structural MRI in medial basal and lateral temporal lobes, and medial and lateral parietal cortex. Furthermore, studies have shown that structural scans identify atrophy patters which distinguish AD from normal aging and other dementias. In particular medial temporal lobe atrophy is associated to impairment in memory, progression from MCI to AD, and can predict future decline and conversion.

To address the problem of early AD diagnosis in living patients, the neuroimaging community is faced with the challenge of improving upon “volumetric measures or visual rating”, “regional patters of atrophy” to satisfy the “more testing is necessary” qualification.

1.2 Morphometry

These notions are made more precise through quantitative brain morphometry at the voxel (mm) scale. Ashburner and Friston [9] describe a partition of morphometry in brain imaging into differences in brain shape (deformation based [DBM] or tensor based morphometry [TBM]) versus differences in local composition of brain tissue after macroscopic differences have been discounted (voxel based morphometry [VBM] i.e. local gray matter concentration).

These techniques require mappings to a standard space often called geometric normalization to establish point to point correspondences between a typically well characterized atlas and a target patient image.

Normalization techniques generally involve a deformation model, a regularization (a.k.a. prior, or penalty) cost, and a data attachment (similarity) cost.

These typically involve an initial low dimensional alignment such as

1. Rotation and translation for intra subject variability (6 dimensions)
2. Rotation, translation, and uniform scaling to preserve shape (7 dimensions)
3. Rotation, translation, and on axis scaling (9 dimensions)
4. Full affine transformations (12 dimensions)
5. Normalization to Talairach space [10] (12 blocks each piecewise affine)

These methods are suitable for providing a coarse correspondence between images, but since these linear maps have a spatially constant Jacobian, they cannot be used for TBM or other analyses of spatially varying growth and atrophy.

Local anatomical differences are accounted for through nonlinear warping techniques. There are a multitude of such techniques, many of which are compared in Klein et al. [11] and reproduced here.

In Advanced Image Registration (AIR), Woods et al. [12][13] describe polynomial basis functions used in Automated Image Registration (AIR). The transformation remains fairly low dimensional.

In Statistical Parametric Mapping (SPM), Friston et al. [14] use low frequency sine or cosine basis functions (depending on desired boundary conditions). Matching problem are linearized for fast explicit (noniterative) solution, valid for small deformations

In Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL) with the FMRIB Nonlinear Image Registration Tool (FNIRT), Andersson, Jenkinson, Smith, et al. [15] model displacement with sine/cosine or B spline basis functions in Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL) with the FMRIB Nonlinear Image Registration Tool (FNIRT). Transformation is regularized with “membrane energy” (penalty on sum of squares of components of derivative).

Table 1.1: Deformation model, approximate number of degrees of freedom (dof), similarity measure, and regularization method for each of the algorithms evaluated in this study. MRes=multiresolution; MSD=mean squared difference; SSD=sum of squared differences; CC=cross-correlation; CR=correlation ratio; MI=mutual information

Algorithm	Deformation	≃dof	Similarity	Regularization
FLIRT	Linear, rigid-body	9.6	normalized CR	
AIR	5th-order polynomial warps	168	MSD (optional intensity scaling)	Incremental increase of polynomial order; MRes: sparse-to-fine voxel sampling
ANIMAL	Local translations	69K	CC	MRes, local Gaussian smoothing; stiffness parameter weights mean deformation vector at each node
ART	Non-parametric, homeomorphic	7 M	normalized CC	MRes median and low-pass Gaussian filtering
Diffeomorphic Demons	Non-parametric, diffeomorphic	21 M	SSD	MRes: Gaussian smoothing
FNIRT	Cubic B-splines	30 K	SSD	Membrane energy; number of basis components; MRes: down-to-up-sampling
IRTK	Cubic B-splines	1.4 M	normalized MI	None used in the study; MRes: control mesh spacing and Gaussian smoothing
JRD-fluid	Viscous fluid: variational calculus (diffeomorphic)	2 M	Jensen-Rényi divergence	Compressible viscous fluid governed by the Navier-Stokes equation for conservation of momentum; MRes

Algorithm	Deformation	≃dof	Similarity	Regularization
ROMEO	Local affine (12 dof)	2 M	Displaced frame difference	First-order explicit regularization method, brightness constancy constraint; MRes: adaptive multigrid (octree subdivision), Gaussian smoothing
SICLE	3-D Fourier series (diffeomorphic)	8 K	SSD	Small-deformation linear elasticity, inverse consistency; MRes: number of basis components
SyN	Bi-directional diffeomorphism	28 M	CC	MRes Gaussian smoothing of the velocity field; transformation symmetry
SPM5:				
"SPM2-type" Normalization	Discrete cosine transforms	1 K	MSD	Bending energy, basis cutoff
Normalization	Discrete cosine transforms	1 K	Generative segmentation model	Bending energy, basis cutoff
Unified Segmentation	Discrete cosine transforms	1 K	Generative segmentation model	Bending energy, basis cutoff
DARTEL Toolbox	Finite difference model of a velocity field (constant over time, diffeomorphic)	6.4 M	Multinomial model ("congealing")	Linear-elasticity; MRes: full multigrid (recursive)

FreeSurfer uses an approach for displacement without basis functions. Vector field with regularization is described by Fischl et al. [16] and has three regularization terms:

- Topology term to penalize small Jacobians
- Metric term to keep neighboring voxels near one another, possibly only for neighbors in the same tissue type
- Smoothness term to keep a voxel close to the average of its neighbors

The data attachment term based on physical model of MRI signal for different tissue types.

For the Demons algorithm, Thirion [17] combines small transformations under composition, rather than vector addition. Small updates are calculated as a smoothed “force term”. Data attachment is based on polarity (inside/outside) at various sites in the image.

In the same vein, Analysis of Functional NeuroImages (AFNI) uses composition at different scales for 3DQWarp [18]. 5th order Hermite polynomials at global scale, composed with 3rd order polynomials on overlapping patches at shrinking sizes. “Simple” constraints on polynomial coefficients keep invertibility at each scale.

It is evident that an immense variety of such mapping algorithms have emerged from the literature. While they are quite heterogeneous, some common themes have emerged to enforce invertibility and regularity of the mapping; they must be diffeomorphic as illustrated in Fig. 1.1. In some cases constraints are used. In others smoothness is enforced through basis functions, or regularization. The notion of using compositions of small vector fields, rather than one large vector field, to model displacement is often used.

1.3 Diffeomorphism framework

These ideas are made more precise by Miller, Trouvé, and Younes [19] by modeling mappings, $\varphi : \Omega \subset \mathbb{R}^3 \to \Omega$ (meaning the mapping φ is a function with domain a subset of $\mathbb{R}^3$ and range the same subset) as flows of the ordinary differential equation (ODE)

$$\frac{d}{dt}\varphi_t = v_t \circ \varphi_t$$

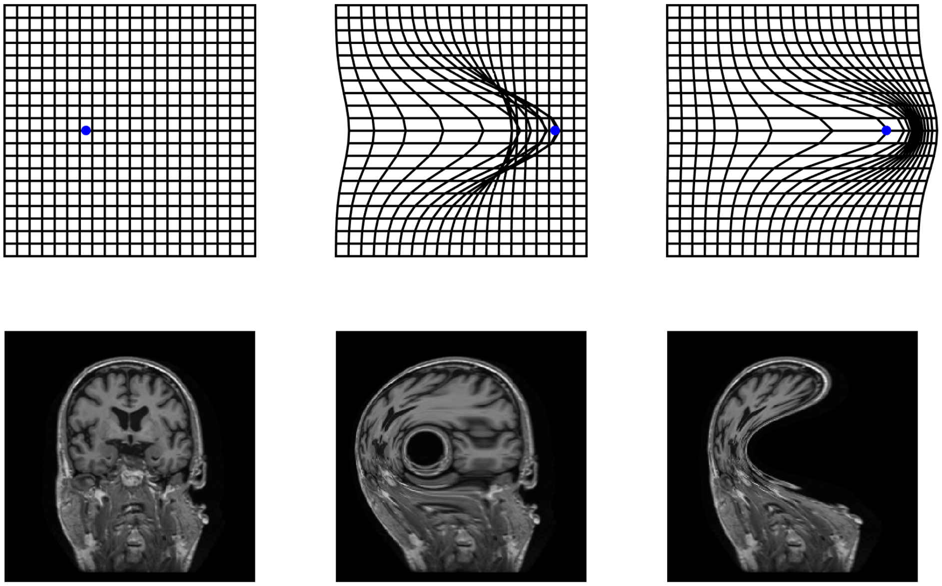


Figure 1.1: Top, a deformed grid illustrating a transformation φ. Bottom a sagittal brain image deformed with $I \circ \varphi$. Left: $\varphi = \mathrm{Id}$, middle φ smooth but not diffeomorphic, right φ a diffeomorphism. Note that the middle and the right columns both result in the same displacement of the blue landmark point. Note that this image is not deformed with the φ^{-1}, which does not exist in the center column.

where $v_t \in V$ where Dupuis, Grenander, and Miller [20] shows we need $V \subset C_0^1(\Omega, \mathbb{R}^3)$ (continuous functions with one continuous derivative, vanishing on the boundary of Ω taking values in $\mathbb{R}^3$). This inclusion occurs when $\|v_t\|_V \leq C\|v_t\|_{1,\infty}$ (the norm in V is less than or equal to some constant times the maximum absolute magnitude of v_t and its first derivatives).

Sobolev inclusion theorems describe how we can generate this space using weakly continuous functions: V can be a Hilbert Sobolev space $H^k(\Omega, \mathbb{R}^3)$ (functions from Ω to $\mathbb{R}^3$ which have a finite modulus square of their first k derivatives) if $k > \frac{5}{2}$. We build functions in this space by defining an inner product through an operator $L : \langle v, w\rangle_V = \langle Lv, Lw\rangle_{L_2}$ with sufficiently many derivatives. Equivalently, we can use a reproducing kernel Hilbert space approach with kernel K, and $\langle K(\cdot, x), K(\cdot, y)\rangle_V = K(x, y)$. In all the work that follows, K will be taken as a translation and rotation invariant Gaussian, which has an infinite number of continuous derivatives.

Using this model to determine point to point correspondences between images was first performed by Beg et al. [21].

Problem 1 (Image matching with flows of diffeomorphisms)**.** *To identify the optimal diffeomorphic mapping between template image I and target image J, we minimize a cost function given by the sum of a regularization term and an error term*

$$\arg\min_{v_t \in V} \int_0^1 \|v_t\|_V^2 dt + \frac{1}{\sigma^2}\|\varphi_1 \cdot I - J\|_{L_2}^2$$

with $\frac{d}{dt}\varphi = v_t(\varphi_t)$ *and* $\varphi_0 = \mathrm{Id}$, *and* $\varphi \cdot I = I \circ \varphi^{-1}$.

This problem is solved in Appendix A, giving the necessary condition $\frac{1}{\sigma_R^2}(L^*L)v_t(x) - \frac{1}{\sigma_M^2} DI_t^T(x)(I_0 \circ \varphi_t^{-1} - J \circ \varphi_{1t}^{-1}) = 0$, where $\varphi_{1t} \doteq \varphi_t \circ \varphi_1^{-1}$.

Because the regularization term involves an integral over velocity squared, any solution must be a length minimizing path, or geodesic. Geodesic trajectories in this context are characterized in Appendix B. Importantly, the entire time varying trajectory φ_t can be characterized by the initial condition to a geodesic equation, v_0.

Definition 1. [Geodesic coordinates] When using a geodesic equation to construct diffeomorphisms from initial conditions, we refer to the vector field v_0 as the "geodesic coordinates" of the diffeomor-

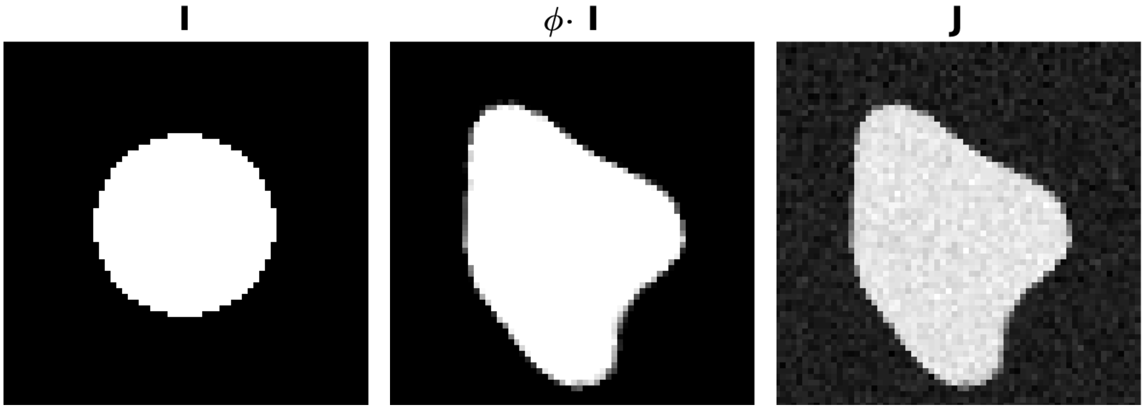


Figure 1.2: The random orbit model is illustrated in three figures. The atlas image (left) is deformed with φ (middle), drawn from some prior distribution. Conditioned on this deformed atlas, the observed imagery J is corrupted by noise.

phism.

It is worth noting that there is an asymmetry in the data attachment proposed here, the template and target image are treated differently. The Symmetric Normalization (SyN) tool in Advanced Normalization Tools (ANTs), created by Avants et al. [22], uses this framework for generating flows of diffeomorphisms, but defines two

$$\varphi_t = \mathrm{Id} + \int_0^t v_s(\varphi_s)dx, \qquad \text{and} \qquad \psi_t = x + \int_1^t v_s(\psi_s)ds$$

one starting at the atlas (time 0), and one starting at the target (time 1). The minimize the cost function

$$\arg\min_{v_t \in V} \int_0^1 \|v_t\|^2 dt + \frac{1}{\sigma^2}\|\varphi_{0.5}\cdot I - \psi_{0.5}\cdot J\|_{L_2}^2$$

This decision to treat the two images identically is a design choice, useful in some situations.

The motivation for Beg et al.'s choice comes from a probabilistic representation of medical images called the random orbit model, in which the observed imagery J is modeled as a conditionally Gaussian random field with mean field $I \circ \varphi^{-1}$. This model is illustrated in Fig. 1.2. Note that the atlas I and target J are of a fundamentally different nature.

The appeal of this representation is that minimizing the cost function can now be interpreted as maximizing a log likelihood. Beg et al.'s cost function corresponds to J as a conditionally Gaussian random field with mean $\varphi \cdot I$ and a residual independent at each voxel with variance σ^2. This probabilistic representation will be key for designing efficient representations of our diffeomorphisms.

1.4 General methods

In general we will approach diffeomorphic mapping as minimum energy problems, defining appropriate regularization and data fidelity costs, and computing the diffeomorphic transformations that will minimize them. We describe these problems through a state, consisting of surfaces and or images that we wish to match to a target; and the transformation being applied to it, a diffeomorphism, the velocity field that generates it, and the parameterization of this velocity field. The relationships between these variables will be enforced by including Lagrange multipliers in an augmented cost function. First order perturbations are considered with respect to each variable, and necessary conditions are derived that characterize stationary solutions. These conditions are used to implement gradient descent algorithms for iteratively computing optimal mappings.

1.5 Outline of the thesis

The thesis is organized as follows. First, we perform a reduction of complexity on v_0, representing it through a singular distribution supported on anatomical boundaries that we call singular geodesic coordinates.

Next, we model probabilistic distribution of these coordinates through a low dimensional parameterization, the geodesic parameters, which allows us to perform filtering and denoising of noisy imagery.

We briefly discuss statistical analysis performed along side diffeomorphic mapping, testing hypotheses about populations of anatomy.

Next we discuss more appropriate data attachment cost functions for matching labeled images, and study a population of amnestic MCI patients.

Finally we develop more advanced deformation models for estimating population templates and deforming them templates onto timeseries of data.

Chapter 2

Flows of diffeomorphisms

2.1 Singular initial momentum representation

Miller, Trouvé, and Younes [23] describe how specific forms of geodesic coordinates are optimal for specific data fidelity cost functions. These descriptions are conveniently represented in terms of the quantity L^*Lv_0 (* referring to the adjoint), which we call the initial momentum. For image matching, the optimal initial momentum is proportional to the image gradient as discussed in Appendix A. For matching of corresponding landmark points, the optimal initial momentum is singularly supported on the points.

For surface matching the optimal initial momentum is supported singularly on the surface. The reduced complexity of this representation has led to robustness of mappings in the presence of noisy or anomalous data. As well, the reduced representation has been important for multiple comparison correction in statistical hypothesis testing. For these reasons we will use the singular surface representation as an effective model for anatomical variation, even in cases where it may not be exactly optimal (such as grayscale image matching).

We assume that the boundaries of anatomical structures we wish to model can be represented by surfaces using the parameterization $f : U \subset \mathbb{R}^2 \to \mathbb{R}^3$, which is indexed over time as it moves with the flow according to $f_t = \varphi_t(f_0)$ for $t \in [0, 1]$ (with the exception of Sec. 9). We introduce the function $p : U \to \mathbb{R}^3$ and represent our velocity as

$$v(x) = \int K(x, f(u))p(u)du \tag{2.1}$$

for some kernel K, chosen such that is inverse to L in the sense that $(L^*L)K(\cdot, x) = \delta(\cdot - x)$ (the Dirac measure). In this case we have

$$\|v\|_V^2 = \iint_{U\times U} p^T(u)K(f(u), f(u'))p(u')dudu'$$

In this work we only consider scalar valued (as opposed to matrix valued) kernels.

Dynamically, our surface parameterization will deform with the flow according to

$$\frac{d}{dt}f = v(f) = K_{ff}p \tag{2.2}$$

where

$$K_{ff}p(u) \doteq \int_U K(f(u), f(u'))p(u')du'$$

defining K_{ff} as the operator that takes a function on U as an input, and outputs another function on U through the integral above.

Because these kernels are related to inner products, they must be symmetric, $K(x, y) = K(y, x)$. In what follows we choose a kernel to be stationary, $K(x, y) = K(x - y, 0)$, which allows for some simplifications in terms of derivatives with respect to each argument. See section C.6.

2.1.1 Singular geodesic coordinates

Statement 1 (Geodesic equation for surface parametrizations)**.** We can characterize geodesics by minimizing the cost

$$\int_0^1 \frac{1}{2} \iint_{U\times U} p^T(u)K(f(u), f(u'))p(u')dudu'dt$$

with fixed endpoints f_0 and f_1. These geodesics are described by the equation

$$\frac{d}{dt}p(u) = -Dv^T(f(u))p(u) \tag{2.3}$$

Proof. We enforce dynamical constraints using Lagrange multipliers, resulting in the augmented cost

$$\int_0^1 \frac{1}{2}\iint_{U\times U} p^T(u)K(f(u),f(u'))p(u')dudu' + (\lambda^f|\frac{d}{dt}f(\cdot) - \int_U K(f(\cdot),f(u))p(u)du)dt$$

where we are using the notation for natural pairing $(\cdot|\cdot)$ to mean transposing the left argument and integrating over the appropriate domain (here U).

We first take a variation with respect to p

$$\begin{aligned}
&\frac{d}{d\epsilon}\int_0^1 \frac{1}{2}\iint_{U\times U} (p(u)+\epsilon\delta p(u))^T(u)K(f(u),f(u'))(p(u')+\epsilon\delta p(u'))dudu' \\
&\qquad + (\lambda|\frac{d}{dt}f(u) - \int_U K(f(u),f(u'))(p(u')+\epsilon\delta p(u'))du')dt\Big|_{\epsilon=0} \\
&= \int_0^1 \iint_{U\times U} p(u)^T K(f(u),f(u'))\delta p(u')dudu' - (\lambda(u)|\int_U K(f(u),f(u'))\delta p(u')du')dt \\
&= \int_0^1 (v(u) - K_{ff}\lambda|\delta p(u))dt
\end{aligned}$$

So we have $v(u) = \int_U K(f(u),f(u'))\lambda(u')du'$, which implies that $\lambda = p$.

Next we extremize over the state f.

$$\begin{aligned}
&\frac{d}{d\epsilon}\frac{1}{2}\iint p^T(u)K(f(u)+\epsilon\delta f(u), f(u')+\epsilon\delta f(u'))p(u')dudu' \\
&\qquad + \int_U \lambda^T(u)\frac{d}{dt}\delta f(u)du - \iint_{U\times U} \lambda^T(u)K(f(u)+\epsilon\delta(u), f(u')+\epsilon\delta(u'))p(u')dudu'|_{\epsilon=0} \\
&= \iint p^T(u)D_1K(f(u),f(u'))\delta f(u)p(u')dudu' - \int_U \frac{d}{dt}\lambda^T(u)\delta f(u)du \\
&\qquad - \iint_{U\times U} \lambda^T(u)D_1K(f(u),f(u'))\delta f(u)p(u')dudu' - \iint_{U\times U} \lambda^T(u)D_2K(f(u),f(u'))\delta f(u')p(u')dudu' \\
&= \iint_{U\times U} p^T(u)p(u')D_1K(f(u),f(u'))\delta f(u)dudu' - \int_U \frac{d}{dt}\lambda^T(u)\delta f(u) \\
&\qquad - \iint_{U\times U} \lambda^T(u)p(u')D_1K(f(u),f(u'))\delta f(u)dudu' - \iint_{U\times U} \lambda^T(u')p(u)D_2K(f(u'),f(u))\delta f(u)dudu'
\end{aligned}$$

Note that because kernel functions are symmetric, $K(x,y) = K(y,x)$, we can write $D_1K(x,y) = D_2K(y,x)$ as shown in (C.7).

This implies that the three terms are in fact all the same, and we are left with

$$(-\int_U DK^T(f(u),f(u'))p^T(u')p(u)du' - \frac{d}{dt}p(u)|\delta f(u))$$

This gives the geodesic equation

$$\frac{d}{dt}p(u) = -\int_U D_1 K^T(f(u), f(u'))p^T(u')p(u)du'$$

Note that this can be written as

$$\frac{d}{dt}p(u) = -Dv^T(f(u))p(u) \tag{2.4}$$

One can readily observe that this equation the same as (B.2), restricted to the deforming surface f. $\square$

Definition 2. [Singular Geodesic Coordinates] When using the geodesic equation to construct diffeomorphisms from initial conditions with the singular representation described above, we refer to the function p_0 as the "singular geodesic coordinates" of the diffeomorphism.

2.1.2 More general parameterizations

Note that in general we deal with surfaces that are not diffeomorphic to a subset of $\mathbb{R}^2$ (such as closed surfaces, several separate surfaces, or discrete surfaces). To be precise in these situations we can introduce U as a measure space with measure η. This space can be a sphere, supporting closed surfaces, or a union of spheres, supporting several closed surfaces, or a set of points with counting measure, supporting discrete surfaces. However, for simplicity we will maintain the Reimannian integral notation that is familiar to most readers. It should be understood that with this convention we have (for example)

$$\int_U K(f(u), f(u))p(u)du \doteq \int_U K(f(u), f(u'))d\eta(u).$$

Chapter 3

Surface matching with singular geodesic coordinates

Surface matching was originally formulated by Vaillant and Glaunès [24]. Our goal will be to parametrically embed the surface description in lower dimensional representations. Therefore it is convenient to reinterpret their original formulation in terms of the singular geodesic coordinates indexed over the surfaces. This work was published as part of [25]. Because we will build building on this algorithm to construct a more parsimonious parameterization of diffeomorphisms, we present a derivation here.

3.1 Cost function and dynamical constraints

We formulate the matching as a constrained optimization problem.

Problem 2 (Surface matching with singular geodesic coordinates)**.** *Given a template surface parameterization f_0, some data fidelity cost F (e.g. current matching as described by Vaillant and Glaunès [24] or varifold matching as described by Charon and Trouvé [26]), and some regularization cost R (e.g. norm squared of v_0) we seek to minimize the cost*

$$C(p_0) = R(p_0) + F(f_1)$$

over the singular geodesic coordinates p_0*, subject to the dynamical constraints* $\frac{d}{dt}f = K_{ff}p = v(f)$ *and* $\frac{d}{dt}p = -Dv^T(f)p$*.*

We begin by incorporating the constraints by writing an augmented cost function in terms of Lagrange multipliers λ^f and λ^p.

$$R(p_0) + F(f_1) + \int_0^1 (\lambda^p|\frac{d}{dt}f - v(f)) + (\lambda^p|\frac{d}{dt}p + Dv^T(f)p)dt \tag{3.1}$$

$$= R(p_0) + F(f_1) + \int_0^1 (\lambda^f|\frac{d}{dt}f - \int_U K(f(u), f(u'))p(u')du') \tag{3.2}$$

$$+ (\lambda^p|\frac{d}{dt}p + \int_U \partial_1 K^T(f(u), f(u'))p(u')^T p(u)du')dt \tag{3.3}$$

Note that generally R may depend on f_0. Since this is fixed, we suppress writing f_0 as an argument.

3.2 Necessary conditions

Statement 2 (Necessary conditions for surface matching with singular geodesic coordinates)**.** For stationary solutions to Problem 2 it is necessary that the λ satisfy the following dynamical equations:

$$\begin{aligned}\frac{d}{dt}\lambda^f &= \int -[\lambda^{fT}(u)p(u') + \lambda^{fT}(u')p(u)]D_1K^T(f(u), f(u'))du' \\ &\quad + \int p^T(u')p(u)[D_1\nabla_1 K(f(u), f(u'))\lambda^p(u) + D_2\nabla_1 K(f(u), f(u'))\lambda^p(u')]du' \end{aligned}\tag{3.4}$$

$$\frac{d}{dt}\lambda^p = -\int K(f(u), f(u'))\lambda^f(u')du' + \int p(u')[D_2K(f(u), f(u'))\lambda^p(u') + D_1K(f(u), f(u'))\lambda^p(u)]du' \tag{3.5}$$

or for stationary kernels

$$\begin{aligned}\frac{d}{dt}\lambda^f &= -\int [\lambda^{fT}(u)p(u') + \lambda^{fT}(u')p(u)]D_1K^T(f(u), f(u'))du' \\ &\quad + \int (p^T(u)p(u')D_1\nabla_1 K(f(u), f(u'))[\lambda^p(u) - \lambda^p(u')]du' \\ \frac{d}{dt}\lambda^p &= -\int K(f(u), f(u'))\lambda^f(u')du' + \int p(u')D_1K(f(u), f(u'))[\lambda_p(u) - \lambda_p(u')]du'\end{aligned}$$

and the boundary conditions $\lambda^p_1 = 0$ and $\lambda^p_0 = DR^T(p_0)$ and $\lambda^f_1 = -DF^T(f_1)$

Proof. The augmented cost function must be stationary to perturbations in each variable. In what follows, there will be a subsection for perturbations for each variable, and sub sub sections describing each term in detail.

3.2.1 Perturbation with respect to f

We perturb the augmented cost function of (3.1) with respect to f, removing terms that do not vary with the perturbation for compactness.

$$\begin{aligned}&\frac{d}{d\epsilon}F(f_1+\epsilon\delta f_1)\Big|_{\epsilon=0}+\frac{d}{d\epsilon}\int_0^1(\lambda^f|\frac{d}{dt}f+\epsilon\frac{d}{dt}\delta f)dt\Big|_{\epsilon=0}\\ &\quad-\frac{d}{d\epsilon}\int_0^1(\lambda^f|\int K(f(u)+\epsilon\delta f(u),f(u')+\epsilon\delta f(u'))p(u')du')dt\Big|_{\epsilon=0}\\ &\quad-\frac{d}{d\epsilon}\int_0^1(\lambda^p|-\int D_1K^T(f(u)+\epsilon\delta f(u),f(u')+\epsilon\delta f(u'))p^T(u')p(u)du')dt\Big|_{\epsilon=0}\end{aligned}$$

Perturbation of boundary terms, and $\frac{d}{dt}f$ with respect to f

Examining these terms gives

$$(DF^T(f_1)|\delta f_1)+\int_0^1(\lambda^f|\frac{d}{dt}\delta f)dt=(DF(f_1)|\delta f_1)+\int_0^1(-\frac{d}{dt}\lambda^f|\delta f)dt+(\lambda_1^f|\delta f_1)$$

after applying integration by parts and recalling $\delta f_0=0$.

For a stationary solution, the boundary term implies that $\lambda_1^f=-DF^T(f_1)$. The term $\frac{d}{dt}\lambda^f$ will be balanced against those below to achieve stationarity.

Perturbation of f dynamics with respect to f

We suppress the integral over time here.

$$\begin{aligned}&-\iint\lambda^{fT}(u)D_1K(f(u),f(u'))\delta f(u)p(u')dudu'-\iint\lambda^{fT}(u)D_2K(f(u),f(u'))\delta f(u')p(u')dudu'\\ &\quad=-\iint\lambda^{fT}(u)D_1K(f(u),f(u'))\delta f(u)p(u')dudu'-\iint\lambda^{fT}(u')D_2K(f(u'),f(u))\delta f(u)p(u)dudu'\\ &\quad=-\iint\lambda^{fT}(u)D_1K(f(u),f(u'))\delta f(u)p(u')dudu'-\iint\lambda^{fT}(u')D_1K(f(u),f(u'))\delta f(u)p(u)dudu'\\ &\quad=(-\int[\lambda_f^T(u)p(u')+\lambda_f^T(u')p(u)]D_1K^T(f(u),f(u'))du'|\delta f(u))\end{aligned}$$

Where we changed the dummy variables and and then used kernel symmetry (C.7) in the right hand term.

Perturbation of p dynamics with respect to f

$$\iint\lambda^{pT}(u)\nabla_1D_1K(f(u),f(u'))\delta f(u)p^T(u')p(u)dudu'$$

$$
\begin{aligned}
&+ \iint \lambda^{pT}(u)\nabla_1 D_2 K(f(u), f(u'))\delta f(u')p^T(u')p(u)dudu' \\
&= \iint \lambda^{pT}(u)\nabla_1 D_1 K(f(u), f(u'))\delta f(u)p^T(u')p(u)dudu' \\
&+ \iint \lambda^{pT}(u')\nabla_1 D_2 K(f(u'), f(u))\delta f(u)p^T(u)p(u')dudu' \\
&= \iint \lambda^{pT}(u)\nabla_1 D_1 K(f(u), f(u'))\delta f(u)p^T(u')p(u)dudu' \\
&+ \iint \lambda^{pT}(u')\nabla_2 D_1 K(f(u), f(u'))\delta f(u)p^T(u)p(u')dudu' \\
&= (\int p^T(u')p(u)[D_1\nabla_1 K(f(u), f(u'))\lambda^p(u) + D_2\nabla_1 K(f(u), f(u'))\lambda^p(u')]du'|\delta f(u))
\end{aligned}
$$

As above, we first change dummy variables, then apply kernel symmetry.

For stationary kernels we can simplify further using (C.8)

$$
(\int p^T(u)p(u')D_1\nabla_1 K(f(u), f(u'))[\lambda_p(u) - \lambda_p(u')]du'|\delta f(u))
$$

Determining λ_f dynamics

We can combine these terms to give

$$
\begin{aligned}
\frac{d}{dt}\lambda_f = &- \int [\lambda^{fT}(u)p(u') + \lambda^{fT}(u')p(u)]D_1 K^T(f(u), f(u'))du' \\
&+ \int p^T(u')p(u)[D_1\nabla_1 K(f(u), f(u'))\lambda^p(u) + D_2\nabla_1 K(f(u), f(u'))\lambda^p(u')]du'
\end{aligned}
$$

Or for stationary kernels

$$
\begin{aligned}
\frac{d}{dt}\lambda_f = &- \int [\lambda^{fT}(u)p(u') + \lambda^{fT}(u')p(u)]D_1 K^T(f(u), f(u'))du' \\
&+ \int [p^T(u)p(u')D_1\nabla_1 K(f(u), f(u')][\lambda^p(u) - \lambda^p(u')]du'
\end{aligned}
$$

3.2.2 Perturbation with respect to p

We perturb the augmented cost (3.1) with respect to p, removing terms that do not vary with the perturbation.

$$
\begin{aligned}
&\frac{d}{d\epsilon}R(p_0 + \epsilon\delta p_0)\Big|_{\epsilon=0} + \frac{d}{d\epsilon}\int (\lambda^p|\frac{d}{dt}p + \epsilon\frac{d}{dt}\delta p)dt\Big|_{\epsilon=0} \\
&\quad - \frac{d}{d\epsilon}\int (\lambda^f|\int K(f(u), f(u'))(p(u') + \epsilon\delta p(u'))du')dt\Big|_{\epsilon=0}
\end{aligned}
$$

$$-\frac{d}{d\epsilon}\int(\lambda^p|-\int\nabla_1 K(f(u),f(u'))(p^T(u)+\epsilon\delta p^T(u))(p(u')+\epsilon\delta p(u'))du')dt\Big|_{\epsilon=0}$$

Perturbation of boundary terms and $\frac{d}{dt}p$ with respect to p

Examining these terms gives

$$(DR^T(p_0)|\delta p_0)+\int(\lambda^p|\frac{d}{dt}\delta p)dt=(DR^T(p_0)|\delta p_0)-\int(\frac{d}{dt}\lambda^p|\delta p)dt+(\lambda^p_1|\delta p_1)-(\lambda^p)_0|\delta p_0)$$

after applying integration by parts. This implies the boundary conditions $\lambda^p_1=0$ and $\lambda^p_0=DR^T(p_0)$.

Perturbation of f dynamics with respect to p

$$\begin{aligned}
-(\lambda^f|&\int K(f(u),f(u'))\delta p(u')du')\\
&=-\iint K(f(u),f(u'))\lambda^{fT}(u)\delta p(u')dudu'\\
&=-\iint K(f(u'),f(u))\lambda^{fT}(u')\delta p(u)dudu'\\
&=-\iint K(f(u),f(u'))\lambda^{fT}(u')\delta p(u)dudu'\\
&=(-\int K(f(u),f(u'))\lambda^f(u')du'|\delta p(u))
\end{aligned}$$

where we simply changed the dummy variable and used kernel symmetry.

Perturbation of p dynamics with respect to p

$$\begin{aligned}
&(\lambda^p|\nabla_1 K(f(u),f(u'))p^T(u)\delta p(u'))+(\lambda^p|\nabla_1 K(f(u),f(u'))p^T(u')\delta p(u))\\
&\quad=\iint\lambda^{pT}(u)\nabla_1 K(f(u),f(u'))p^T(u)\delta p(u')dudu'+\iint\lambda^{pT}(u)\nabla_1 K(f(u),f(u'))p^T(u')\delta p(u)dudu'\\
&\quad=\iint\lambda^{pT}(u')\nabla_1 K(f(u'),f(u))p^T(u')\delta p(u)dudu'+\iint\lambda^{pT}(u)\nabla_1 K(f(u),f(u'))p^T(u')\delta p(u)dudu'\\
&\quad=\iint\lambda^{pT}(u')\nabla_2 K(f(u),f(u'))p^T(u')\delta p(u)dudu'+\iint\lambda^{pT}(u)\nabla_1 K(f(u),f(u'))p^T(u')\delta p(u)dudu'\\
&\quad=(\int p(u')[D_2K(f(u),f(u'))\lambda^p(u')+D_1K(f(u),f(u'))\lambda^p(u)]du'|\delta p(u))
\end{aligned}$$

where we change dummy variables, and use kernel symmetry (C.7).

For stationary kernels this gives

$$(\int p(u')D_1K(f(u),f(u'))[\lambda^p(u)-\lambda^p(u')]du'|\delta p(u))$$

Determining λ_p dynamics

Combining these terms gives

$$\frac{d}{dt}\lambda^p = -\int K(f(u), f(u'))\lambda^f(u')du' + \int p(u')[D_2K(f(u), f(u'))\lambda^p(u') + D_1K(f(u), f(u'))\lambda^p(u)]du'$$

Or for stationary kernels

$$\frac{d}{dt}\lambda^p = -\int K(f(u), f(u'))\lambda^f(u')du' + \int p(u')D_1K(f(u), f(u'))[\lambda^p(u) - \lambda^p(u')]du'$$

3.3 Gradient of cost function

Note that we can interpret

$$\frac{\partial}{\partial p_0}F(f_1) = -\lambda_0^p$$

When all the constraints are obeyed (boundary and dynamics) except the one for λ_0^p, the variation in the corresponding parameters is equal to zero. This is the leftover term linear in the perturbation δp_0. In other words it is the derivative of the fidelity term with respect to the the initial momentum.

We can include the regularization term to obtain

$$\frac{\partial}{\partial p_0}C = DR^T(p_0) - \lambda_0^p$$

We can write this differential as a gradient vector using the inner product in V^*

$$\begin{aligned}(DR^T(p_0) - \lambda_0^p|\delta p_0) &= (K_{ff}K_{ff}^{-1}(DR^T(p_0) - \lambda_0^p))|\delta p_0)\\ &= \langle K_{ff}^{-1}(DR^T(p_0) - \lambda_0^p), \delta p_0\rangle_{V^*}\end{aligned}$$

Technically the terms in the inner product above are not elements of V^*. For example $\delta p_0 \notin V^*$, but $\int_U \delta p_0(u)\delta(\cdot - f(u))du \in V^*$, and we identify one with the other for notational convenience.

One special case is worth addressing. Typically $R(p_0) = \frac{1}{2}\iint K(f_0(u), f_0(u'))p^T(u)p(u')dudu'$. In this case the variation $DR(p_0) = \int K(f_0(u), f_0(u'))p_0^T(u')du'$ and so the gradient vector in V^* is simply p_0.

3.4 Algorithm for surface matching with singular geodesic coordinates

This gives the following gradient descent based algorithm for surface matching

Algorithm 3.1. surface matching with singular geodesic coordinates

1. Initialize f_0 to a template, initialize p_0 (typically to 0)
2. Calculate f_1, p_1 by flowing (2.2) and (2.3) forward in time
3. Calculate $DF(f_1)$
4. Initialize $\lambda_1^f = -DF(f_1), \lambda_1^p = 0$
5. Flow λ_f, λ_p backwards to $t = 0$ using (3.4) and (3.5)
6. Calculate gradient vector $K_{ff}^{-1}(DR^T(p_0) - \lambda_{p,0})$
7. Update $p_0 \mapsto p_0 - \epsilon K_{ff}^{-1}(DR^T(p_0) - \lambda_{p,0})$ for some ϵ sufficiently small to reduce the cost
8. Repeat steps 2 to 7 until convergence criteria is met

3.5 Discrete implementation details

Surfaces are discretized using triangulated meshes. The time domain is generally discretized into 10 steps. Dynamical equations for $f, p, \lambda^f, \lambda^p$ are solved using Euler's method, noting that when $\frac{d}{dt}\lambda$ is calculated to update from time $k+1$ to time k, the values of the other variables from time k are used (this is adjoint to a typical Euler step).

Chapter 4

Parametric embedding of surface shapes via geodesic coordinates

Many of the advantages of using a singular representation for geodesic coordinates can be extended by using an even lower dimensional representation. We will choose a parametric parametric representation for p_0 which will be effective for a small number (around 50) parameters.

4.1 Expressing p_0 parametrically

We will express p_0 with a mean function b^0 and linear combination of basis vectors b^i with coefficients β^i for $i \in \{1, \ldots, B\}$.

$$p_0(u) = b^0(u) + \sum_{i=1}^{B} \beta^i b^i(u) \tag{4.1}$$

Definition 3 (Geodesic parameters)**.** When expressing singular geodesic coordinates 2 with respect to a basis, we call the coefficients in the representation the *geodesic parameters*.

When performing matching, we may use our previous regularization function expressed in terms of p_0

$$R(b^0(u) + \sum_{i=1}^{B} \beta^i b^i(u), f_0(u))$$

or we may use some other regularization function $\tilde{R}(\beta)$ expressed in terms of the geodesic parameters directly.

4.2 Gaussian random field model

One important case is when we model the coefficients as multivariate Gaussian with zero mean and covariance Σ. We choose a regularization term as a negative log likelihood.

$$\tilde{R}(\beta) = \frac{1}{2}\beta^T \Sigma^{-1} \beta$$

It follows that we can model p_0 as a Gaussian random field on U with

$$E[p_0(u)] = b^0(u)$$

and

$$\begin{aligned}
\mathrm{Cov}[p_0](u,u') &= E[(p_0(u) - b^0(u))(p_0(u') - b^0(u'))^T] \\
&= E[\sum_{i=1}^{B} \beta^i b^i(u) \sum_{i'=1}^{B} \beta^{i'} b^{i'T}(u')] \\
&= \sum_{i,i'=1}^{B} b^i(u) E[\beta^i \beta^{i'}] b^{i'T}(u') \\
&= \sum_{i,i'=1}^{B} b^i(u) \Sigma^{i,i'} b^{i'T}(u') \\
&\doteq b(u) \Sigma b^T(u')
\end{aligned}$$

where in the last line, summing over basis functions can be thought of as essentially “matrix multiplication”

We can also describe a Gaussian random field on Ω with

$$E[v_0(x)] = \int_U K(x, f_0(u)) b^0(u) du$$

and

$$\mathrm{Cov}[v_0](x,x') = E[\int_U \sum_{i=1}^{B} K(x, f(u)) \beta^i b^i(u) du \int_U \sum_{i'=1}^{B} K(x', f(u')) \beta^{i'} b^{i'T}(u') du']$$

$$
\begin{aligned}
&= \sum_{i,i'=1}^{B} \int K(x,f(u))b^i(u)du E[\beta^i\beta^{i'}] \int_U K(x',f(u'))b^{i'T}(u')du' \\
&= \sum_{i,i'=1}^{B} \int K(x,f(u))b^i(u)du \Sigma^{i,i'} \int_U K(x',f(u'))b^{i'T}(u')du' \\
&\doteq [K_{xf}b]\Sigma[K_{x'f}b]^T
\end{aligned}
$$

where again last line can be thought of as a notation for essentially matrix multiplication.

4.3 Principal component analysis

If enough training data is available, we may choose our low dimensional basis with principal component analysis (PCA). This will provide a model where geodesic parameters are uncorrelated: $\Sigma = \text{diag}[\sigma^2_{\beta^1}, \dots, \sigma^2_{\beta^B}]$, and a negative log likelihood

$$\tilde{R}(\beta) = \sum_{i=1}^{B} \frac{1}{2\sigma^2_{\beta^i}}(\beta^i)^2$$

Further, if we wish to use the interpretation of PCA as the subspace of maximum variance, we will require orthonormality with respect to our inner product in V.

$$\int_{U\times U} b^{iT}(u)K(f(u),f(u'))b^j(u)dudu' = \delta_{ii'}$$

This procedure was proposed and described in detail by Vaillant et al. [27].

We are given a population of initial momentum vectors P, each column being the initial momentum for a single subject, generally discretized with one value per triangulated template surface vertex (with kth vertex at location $f(u^k)$ and M total vertices), with x, y, z components sorted in lexicographic order. We refer to ith observed vector, with jth component (after sorting lexicographically) as P^{ij}

For example, the inner product can be represented through the matrix K with

$$
\begin{aligned}
&\langle b^i, b^j\rangle_{V^*} \\
&= \int_{U\times U} b^{iT}(u)K(f(u),f(u'))b^j(u)dudu' \\
&= \sum_{k,l=1}^{M} b^{iT}(u^k)K(f(u^i),f(u^l))b^j(u^l)
\end{aligned}
$$

$$= \begin{pmatrix} b^{ix}(u_1) \\ b^{iy}(u_1) \\ b^{iz}(u_1) \\ \dots \\ b^{ix}(u_M) \\ b^{iy}(u_M) \\ b^{iz}(u_M) \end{pmatrix}^T \begin{pmatrix} \mathbf{I}K(f(u_1), f(u_1)) & \vdots & \mathbf{I}K(f(u_1), f(u_M)) \\ \vdots & \ddots & \vdots \\ \mathbf{I}K(f(u_M), f(u_1)) & \cdots & \mathbf{I}K(f(u_M), f(u_M)) \end{pmatrix} \begin{pmatrix} b^{jx}(u_1) \\ b^{jy}(u_1) \\ b^{jz}(u_1) \\ \vdots \\ b^{jx}(u_M) \\ b^{jy}(u_M) \\ b^{jz}(u_M) \end{pmatrix}$$

$$\doteq B^i K B^j$$

where $\mathbf{I}$ is the 3×3 identity matrix.

Algorithm 4.1. Principal component analysis of singular geodesic coordinate vectors

1. Given a population of singular geodesic coordinate vectors P, each column being the initial momentum for a single subject, with x, y, z components sorted in lexicographic order.
2. Compute the mean $\bar{P}^i = \frac{1}{N}\sum_{j=1}^{N} P^{ij}$ and centered data $\mathring{P}^{ij} = P^{ij} - \bar{P}^i$
3. Form the Gram matrix $G^{ij} = \langle \mathring{P}^i, \mathring{P}^j \rangle_{V^*}$
4. Compute the eigenvectors of the Gram matrix V and eigenvalues in diagonal matrix D^2.
5. Compute the principal components stored in columns of $U = \mathring{P}V(D^2)^{-1/2}$
6. Compute variance of each component by $\sigma^2_{\beta^i} = \frac{1}{N}D^2_{ii}$
7. Choose a number of dimensions B as the smallest number such that $\sum_{i=1}^{B} \sigma^2_{\beta^i} \geq t \sum_{i=1}^{N} \sigma^2_{\beta^i}$ for some fraction t (for example $t = 0.95$)

4.3.1 Examples

We show several examples computed from the Alzheimer's disease neuroimaging initiative (ADNI) dataset as described in [25]. Template surfaces were generated from published segmentations using restricted Delaunay triangulation [28] and were mapped on to large populations of a variety of

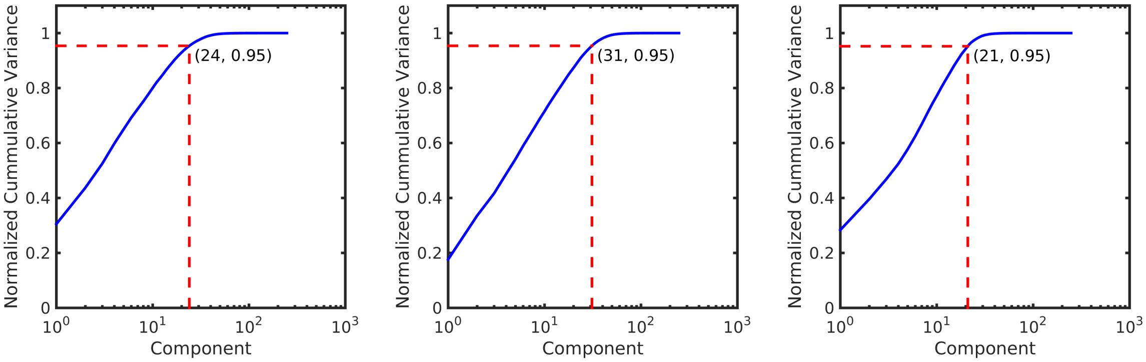


Figure 4.1: Normalized cumulative variance of each coefficient β^i. Most of the population variability is captured by a small number of coefficients. Note log scale on horizontal axis. Three example subcortical structures are shown: left putamen, center hippocampus, right amygdala.

anatomical structures using Algorithm 3.1. The normalized cumulative variance, $\sum_{j=1^i} \sigma^2_{\beta^j} / \sum_{j=1^N} \sigma^2_{\beta^j}$ is shown in Fig. 4.1. The number of components B required for $t = 0.95$ is indicated.

To visualize this coordinate system we can create deformed template surfaces with using geodesic parameters $(\beta^1, \beta^2, 0, \ldots, 0)$. We sample β^1 and β^2 in units of 1.5 standard deviations, and arrange the deformed template surfaces in a grid in Fig. 4.2.

4.3.2 Application to outlier detection

After performing initial momentum surface matching on a population, each structure can be embedded into a low dimensional coordinate system for quantitative or qualitative outlier detection. Structures appearing far from the mean can be easily identified and targeted for quality control. Examples are shown in Fig. 4.3.

Quantitatively, the value of $\tilde{R}(\beta)$ can signal outliers. If the Gaussian model is accurate, $\tilde{R}(\beta)$ is distributed as a χ^2 random variable with N degrees of freedom (a sum of N IID standard normals squared). The χ^2 CDF can be used to convert coordinates to p-values, with small p-values signaling that an observation is unlikely to have come from the Gaussian model. Alternatively, the empirical CDF can be used to convert to p-values, a nonparametric approach, which is shown in Fig. 4.4.

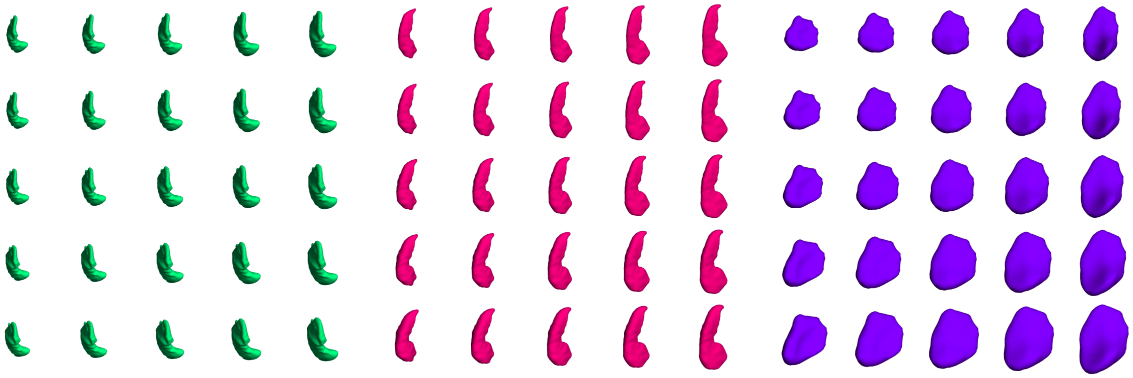


Figure 4.2: Each 5 by 5 grid shows the template surface in the center. Left right corresponds to deforming 1.5 standard deviations the direction of the first mode, whereas up down corresponds to deforming 1.5 standard deviations in the direction of the second mode. Three example subcortical structures are shown: left putamen, center hippocampus, right amygdala.

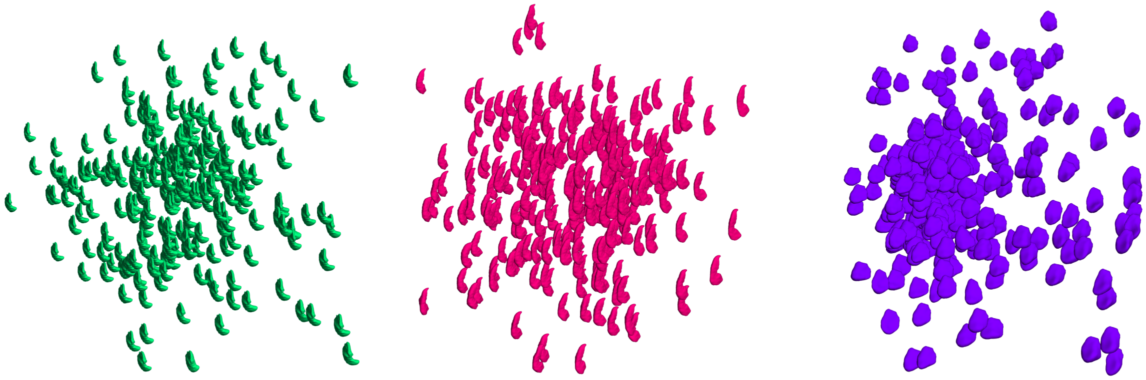


Figure 4.3: Subcortical structures are placed into a 2D coordinate system according to β^1 (horizontal axis) and β^2 (vertical axis). The figure borders correspond to ±3 standard deviations in each direction, and structures with coordinates outside the bounding box are pinned to the border. Three example subcortical structures are shown: left putamen, center hippocampus, right amygdala.

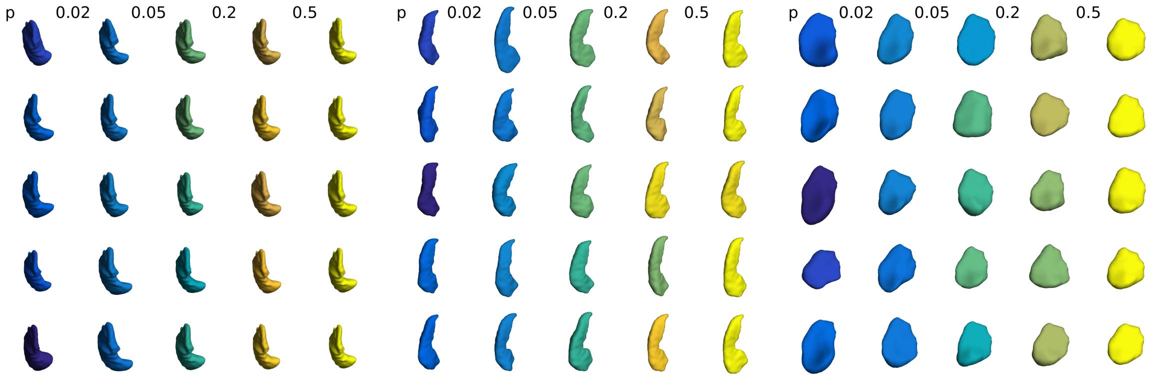


Figure 4.4: Each column shows five example subcortical structures with p value calculated from empirical CDF of $\tilde{R}(\beta)$ lying between the values written between columns. Bright colors show more likely values, whereas dark colors show less likely values. Three example subcortical structures are shown: left putamen, center hippocampus, right amygdala.

4.4 Laplace Beltrami Eigenfunctions

Eigenfunctions of the Surface Laplacian (or Laplace-Beltrami operator) play a role analogous to the Fourier basis in flat space, or the spherical harmonics on a sphere. When training data is unavailable for PCA, these functions can provide a basis to construct geodesic parameters. For discrete triangulate surfaces, the Laplacian operator can be described by the "cotan formula" as described for example by Crane et al. [29]. The discrete Laplacian of a scalar function p defined on the vertices of a triangulated surface $f(u_i)$ is given by

$$(\Delta p)_i = \frac{1}{2}\sum_j (\cot \alpha_j + \cot \beta_j)(p_j - p_i)$$

where we sum over the vertices connected by an edge to vertex i, and α_j and β_j are the two opposite angles to the edge ij in the triangles sharing this edge.

Eigenfunctions are calculated using a standard generalized eigenvalue solver, weighted by vertex area (1/3 of all neighboring triangle areas, which sums correctly to total area), and those corresponding to the smallest eigenvalue (the smoothest or most low frequency functions) are used for a basis.

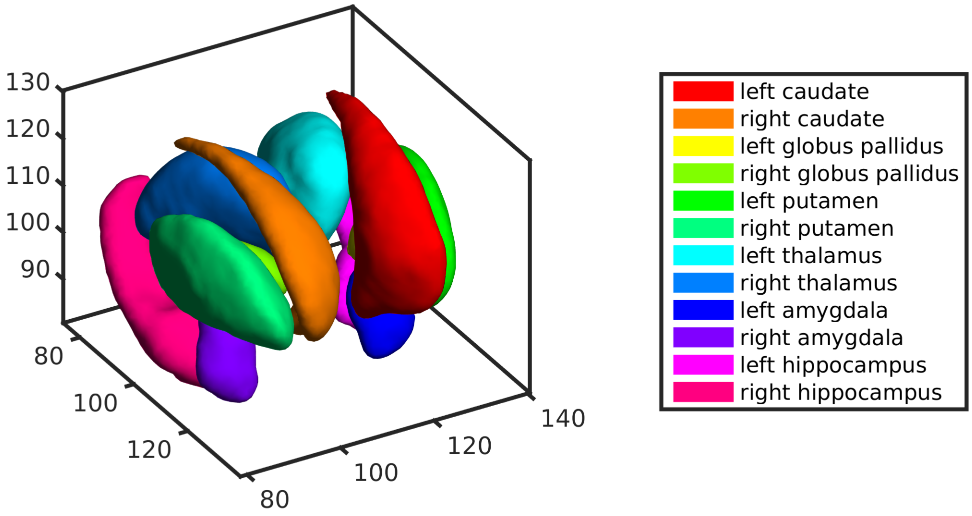


Figure 4.5: Template surfaces on which Laplace Beltrami eigenfunctions are calculated. Axes have units of mm. Structure labels on right.

4.4.1 Examples

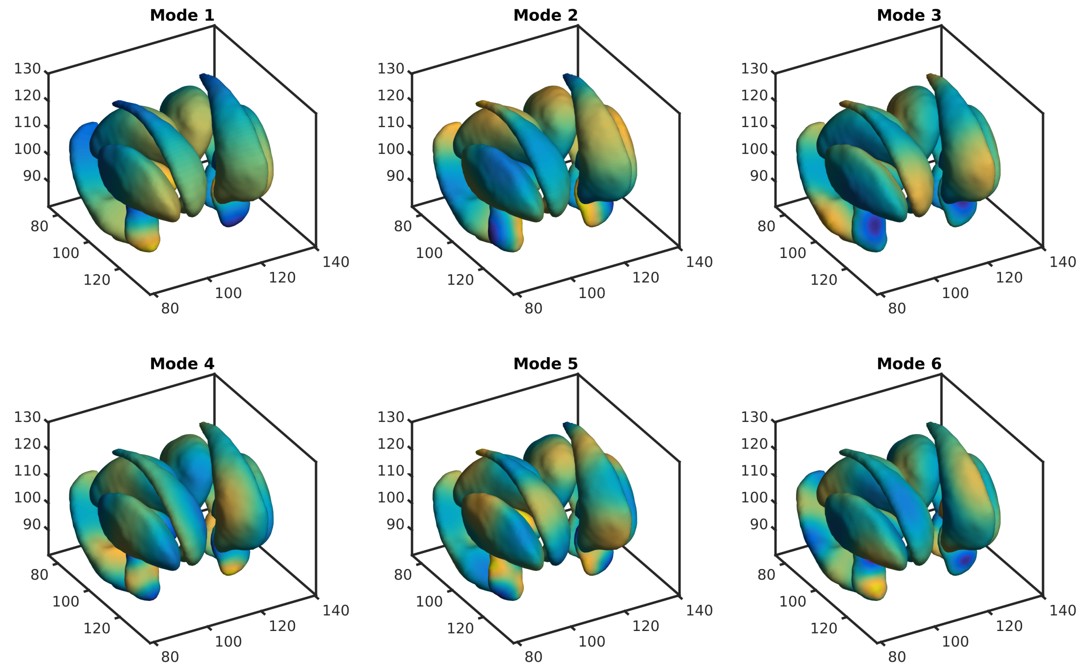
We show examples of Laplace Beltrami eigenfunctions calculated for the surfaces of subcortical structures caudate, globus pallidus, putamen, thalamus, amygdala, and hippocampus. Template surfaces for these examples were generated by restricted Delaunay triangulation [28] of segmentations performed manually in a study related to autism and ADHD [30] and are shown in 4.5.

The first six modes (following the constant functions which have zero eigenvalues) are shown for each structure in Fig. 4.6.

As further illustration, the first two modes are laid out in a 2D coordinate system in Fig 4.7 for three selected structures.

4.5 Surface matching with geodesic parameters

Problem 3 (Surface matching with geodesic parameters)**.** *Given a template surface parameterization f_0, a set of basis functions b^i, some data fidelity cost F, and some regularization cost R, we*


Mode 1

Mode 2

Mode 3

Mode 4

Mode 5

Mode 6

Figure 4.6: The first 6 eigenfunctions of the Laplacian operator (other than the constant functions) are shown for each of the subcortical structures examined. Yellow refers to positive, green to zero, and blue to negative values. Since they are eigenfunctions, their normalization is arbitrary.

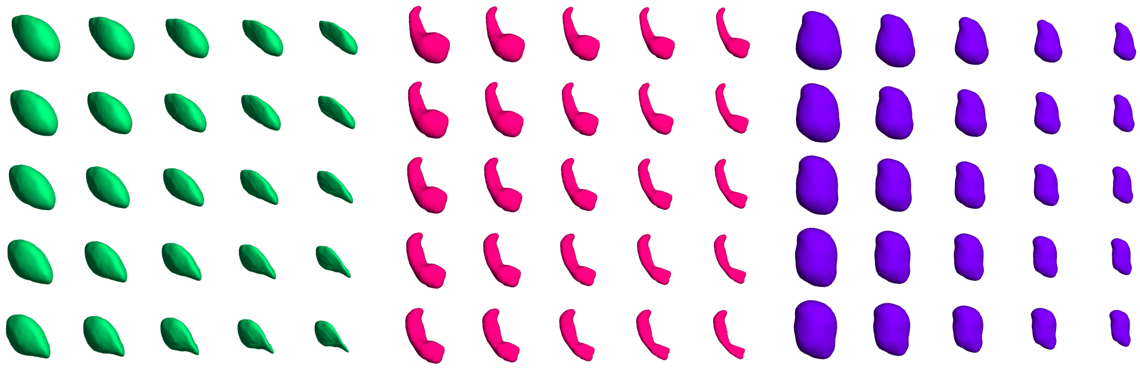


Figure 4.7: Each 5 by 5 grid shows the template surface in the center. Left right corresponds to deforming in the direction of the first (constant) mode, whereas up down corresponds to deforming in the direction of the second mode. Three example subcortical structures are shown: left putamen, center hippocampus, right amygdala.

seek to minimize the cost

$$C(\beta) = R(p_0) + F(f_1) \ \textit{or}\ C(\beta) = \tilde{R}(\beta) + F(f_1)$$

over β*, with geodesic coordinates given by* $p_0 = b^0 + \sum_{i=1}^{B} \beta^i b^i$*, subject to the dynamical constraints* $\frac{d}{dt} f = K_{ff} p = v(f)$ *and* $\frac{d}{dt} p = -Dv^T(f)p$*.*

Statement 3 (Necessary conditions for surface matching with geodesic parameters)**.** Necessary conditions for a stationary solution are given by $\int_U (DR(p_0)(u) - \lambda_0^{pT}(u))b_i(u)du = 0$ when using $R(p_0)$, or $D_i\tilde{R}(\beta) - \int_U \lambda_0^{pT}(u)b_i(u)du$ when using $\tilde{R}(\beta)$.

Proof. We have the simple linear relationship $\delta p_0 = \sum_i b_i \delta\beta_i$, so given the variation

$$\begin{aligned}(DR^T(p_0) - \lambda_0^p | \delta p_0) &= (DR^T(p_0) - \lambda_0^p | \sum_i b_i \delta\beta_i) \\ &= \sum_i \int_U (DR(p_0)(u) - \lambda_0^{pT}(u))b_i(u)\delta\beta_i du\end{aligned}$$

It follows that the gradient with respect to β^i is simply

$$\int_U (DR(p_0)(u) - \lambda_0^{pT}(u))b_i(u)du$$

When using regularization cost $\tilde{R}(\beta)$ instead of $R(p_0)$, we simply have a gradient with respect to β of

$$D_i\tilde{R}(\beta) - \int_U \lambda_0^{pT}(u)b_i(u)du$$

□

This gives the following gradient descent based algorithm for surface matching

Algorithm 4.2. Surface matching with geodesic parameters

1. Initialize f_0 to a template, initialize geodesic parameters β (typically to 0)
2. Compute p_0 from β using (4.1)
3. Calculate f_1, p_1 by flowing (2.2) and (2.3) forward in time
4. Calculate $DF(f_1)$

5. Initialize $\lambda_1^f = -DF^T(f_1), \lambda_1^p = 0$

6. Flow λ^f, λ^p backwards to $t = 0$ using (3.4) and (3.5)

7. Calculate gradient vector $\int_U (DR(p_0) - \lambda_0^{pT}) b_i(u)du$ (or $D_i\tilde{R}(\beta) - \int_U \lambda_0^{pT} b_i(u)du$) for each i

8. Update $\beta_i \mapsto \beta_i - \epsilon \int_U (dR(p_0) - \lambda_0^{pT}) b_i(u)du$ (or $\beta_i \mapsto \beta_i - \epsilon \left(D_i\tilde{R}(\beta) - \int_U \lambda_0^{pT} b_i(u)du\right)$) for some small epsilon

9. Repeat step 2 to 8 until convergence criteria is met.

Algorithm 4.2 is also called geodesic diffeomorphic active shapes (or GDAS) in analogy with Cootes et al. [31]'s description of active shapes.

4.6 Examples

Several examples of surface matching with geodesic parameters are presented below for various fidelity terms F. This data was published in [25] and much of what follows below is reproduced from there.

4.6.1 Current matching

In challenging surface mapping applications it can be necessary to regularize the mappings to avoid undesirable results, and GDAS provides a powerful method for doing so.

In particular, volumetric segmentations of neuroanatomy are often readily available. Converting them to an isosurface for analysis and display is standard, and GDAS provides a method to convert such an isosurface to one reflecting the typicality and variability of a population, rather than features of the volumetric data with an unnatural voxelized structure. Our goal here is to provide a tool to correct for such erroneous segmentations.

This application is essentially equivalent to that presented in [32]. We retain it here as it is the most natural application of GDAS (priors learned from surface matching are used to regularize surface matching), to develop a notation consistent with that for our other applications, and to demonstrate robust performance on poorly behaved datasets.

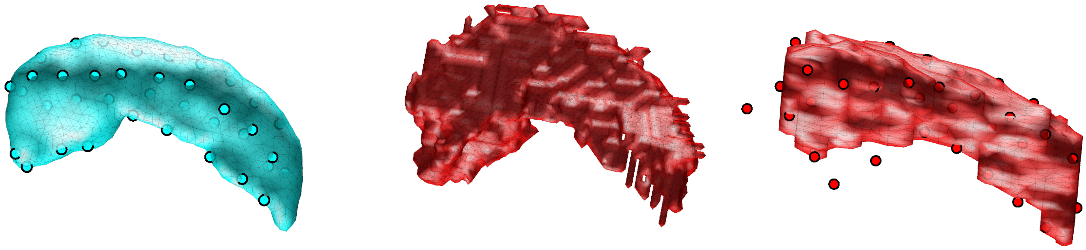

(a) (b) (c)

Figure 4.8: a) Template hippocampus for ADNI dataset. b) Hippocampus isosurface from example volumetric parcellation. c) Isosurface of example hippocampus manual segmentation for our landmark datasets.

Current matching equation brings surfaces f, f^{target} and their unit normals n, n^{target} close

$$\begin{aligned}F(f_1) = &\iint_{U\times U} \exp\left(-\frac{1}{2\sigma_C^2}\left(f_1(u) - f_1(u')\right)^2\right) n_1(u)\cdot n_1(u')du du' \\ &- 2\iint_{U\times U} \exp\left(-\frac{1}{2\sigma_C^2}\left(f_1(u) - f^{\text{target}}(u')\right)^2\right) n_1(u)\cdot n^{\text{target}}(u')du du' \\ &+ \iint_{U\times U} \exp\left(-\frac{1}{2\sigma_C^2}\left(f^{\text{target}}(u) - f^{\text{target}}(u')\right)^2\right) n^{\text{target}}(u)\cdot n^{\text{target}}(u')du du'\end{aligned}$$

where σ_C is a spatial scale parameter defining the size of neighborhood to average over when matching and n is the unit normal to the surface.

The discrete approximation using triangulated surfaces is implemented by letting each triangle face center be written as c^i defined as the average of neighboring vertices, and area weighted face normal N^i being half the cross product of two edge vectors (oriented consistently), we write the discrete current energy as

$$\begin{aligned}F(f_1) = &\sum_i\sum_j \exp\left(-\frac{1}{2\sigma_C^2}\left(c_1^i - c_1^j\right)^2\right) N_1^i\cdot N^j \\ &- 2\sum_i\sum_j \exp\left(-\frac{1}{2\sigma_C^2}\left(c_1^i - c^{j,\text{target}}\right)^2\right) N_1^i\cdot N^{j,\text{target}} \\ &+ \sum_i\sum_j \exp\left(-\frac{1}{2\sigma_C^2}\left(c^{i,\text{target}} - c^{j,\text{target}}\right)^2\right) N^{i,\text{target}}\cdot N^{j,\text{target}}\end{aligned}$$

Methods

As part of the ADNI study, volumetric parcellations (performed using FreeSurfer, described for example in [33]) of whole brains are available at a series of time points. The $t = 0$ data has been studied and a template (see Fig. 4.8a) as well as a population of initial momenta data have been calculated [34]. To study their changing shapes over time, we wish to convert such binary segmentations to surfaces. However, the voxelized nature of the segmentations makes simple isosurfaces unacceptable (as shown in Fig. 4.8b).

We therefore employ the technique of matching our template to such an isosurface, using the constraints of a smooth deformation regularized by PCA to avoid the unnatural appearance of the isosurface. We show examples of GDAS surface matching results, and compare to typical performance using traditional surface matching LDDMM.

Results

For 5 examples, the outcomes of traditional LDDMM surface matching [24] and GDAS surface matching are shown in Fig. 4.9, with target isosurfaces shown on the right hand side. Qualitatively speaking, the traditional LDDMM result tends to produce squared off hippocampal heads (left side in figure) due to outlier voxels, as well as an overestimation of the medial margin (bottom of figure) due to overfitting an outlier "ribbon" of voxels.

The constraints imposed in GDAS surface matching result in a useful and accurate segmentation reflective of the population being analyzed. The "fingerlike" and "ribbonlike" projections reflecting the voxelized structure of the target isosurface, as well as the set of constraints used in FreeSurfer that are designed for an unrelated application, do not significantly influence the resulting surface.

4.6.2 Landmark matching

A further application of our framework involves ROI analysis methods based on diffeomorphic landmark matching. Given a template surface containing K landmarks located on vertices, a trained technician places corresponding landmarks in T1 MR images. Diffeomorphic landmark matching

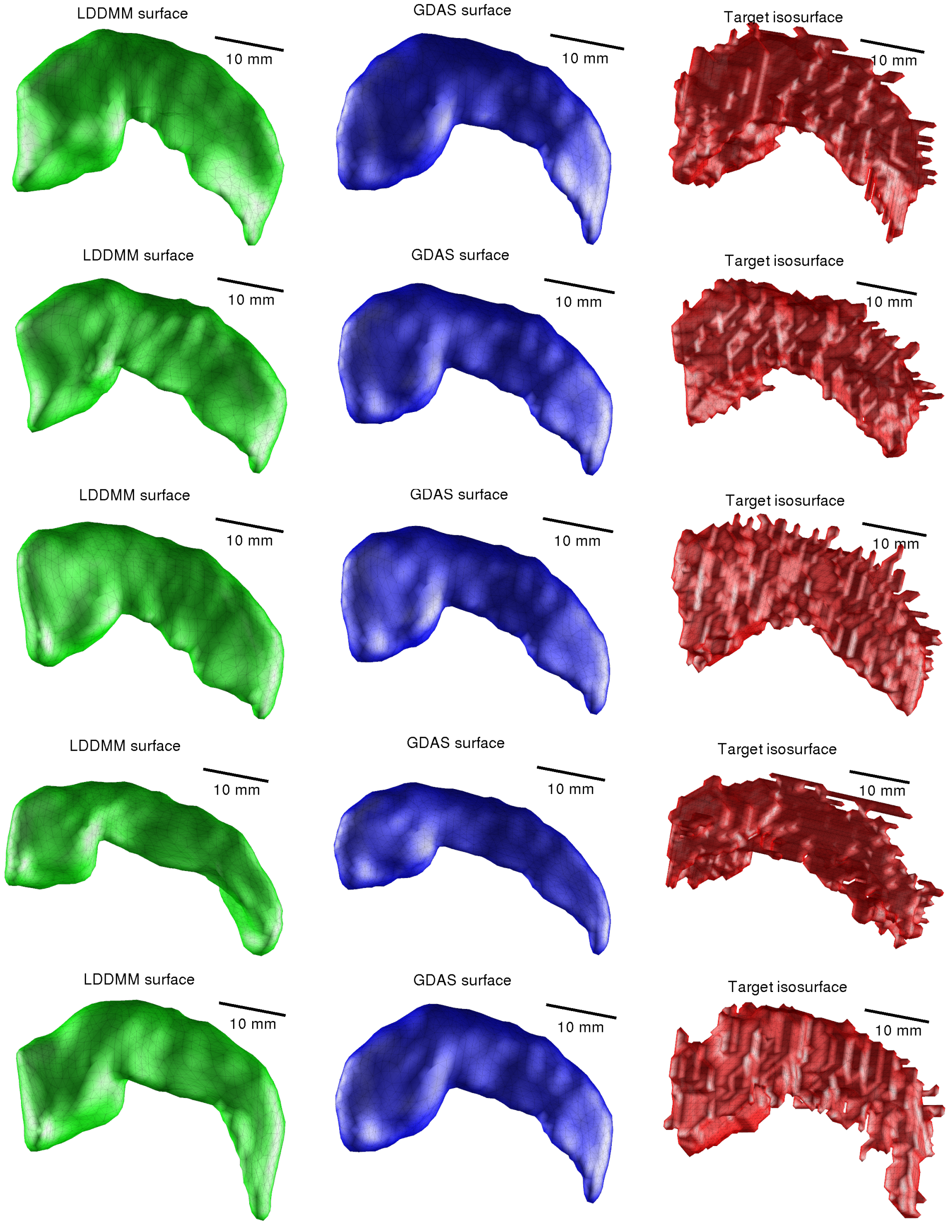


Figure 4.9: Examples of hippocampus surfaces resulting from using (left/green) surface LDDMM, and (center/blue) GDAS surface matching. The target isosurface is shown at right (red).

provides a segmentation of the structure of interest in each T1 image by applying the landmark-based transformation to the entire template surface. This procedure is advantageous, because it provides a compromise between the speed of automatic segmentation, and the accuracy of hand segmentation.

However, variability in landmark placement and sparsity of landmarks can occasionally lead to unsatisfactory segmentation results, and to a time consuming quality control stage where such segmentations are fixed manually. We propose to regularize the problem, taking into account landmark placement variability based on voxel size, as well as shape variability learned from PCA.

For data fidelity term we choose a subset of points in our triangulated surface $f(i)$, and match them to labeled landmarks $l(k)$ with square error penalty.

Methods

Our segmentation pipeline for our ROI methods is described in [35–40]. The relevant portion (the landmark matching phase) for one such study is summarized here. Thirty eight landmarks are placed along the left and right hippocampus in 441 $0.93 \times 0.93 \times 2.0$ mm T1 images of the brain. The first was placed at the tip of the head of the hippocampus (the center of the most anterior slice containing the hippocampus in a T1 image), and the second was placed at the tip of the tail of the hippocampus (the center of the most posterior slice containing the hippocampus). The distance between these two was then divided into 9 slices from anterior to posterior, and on each slice 4 landmarks were placed at the superior, inferior, medial, and lateral margins of the hippocampus. This manual procedure takes approximately 10 minutes for a trained technician to complete, as compared to over 2 hours for a full hand segmentation of images of this size.

In the existing segmentation and analysis pipeline, a template surface was chosen as the left hippocampus for a single subject, and a manual segmentation and resulting isosurface were generated for this case. After a similitude alignment (including reflecting right hippocampi to match left) landmark LDDMM [41] was used to map this template to each target, defining a segmentation surface and binary image for each patient.

However, this procedure was found to suffer from lack of robustness, and roughly 30 out of 441 cases were unacceptable. A laborious phase of quality control was necessary involving identifying problematic or distorted segmentations, manually editing their binary images, and regenerating isosurfaces.

To measure whether our prior model provides enough robustness to avoid such issues, we chose 5 challenging cases of left hippocampi (as identified during quality control inspection), where manual intervention was required, and 11 typical cases, and we examined the performance improvement using the proposed algorithm rather than that outlined above. These cases were manually segmented by a trained technician to provide a gold standard, and associated isosurfaces were generated for further evaluation. Furthermore, for 3 cases requiring intervention and 2 typical cases (those illustrated in Fig. 4.10), a second manual segmentation was obtained to give a sense of inter-rater variability.

Note that the segmentations that are shown here do not constitute the final output of the ROI pipeline described in [35], in which they would be further processed. That is, the results of standard landmark mapping shown here are not reflective of the final segmentations. However, we expect improvement at this stage to contribute to overall improvement.

The template with its associated landmarks is shown in Fig. 4.8a. In Fig. 4.8c, an example isosurface generated from a manual segmentation is shown together with its associated landmarks. Note that landmarks were placed on template surface vertices, but the target landmarks were placed independently (by a different technician) from the gold standard manual segmentation. Fig. 4.8 shows the uncertainty of landmark placement, particularly in the region of the hippocampus' head, and demonstrates the need to include landmark placement variability in the segmentation algorithm. In these figures, and throughout this paper, the color cyan will is used for the template, red for the target, blue for our new results and green for results using existing algorithms.

Results

For 3 cases requiring quality control and 2 typical cases, the outcome of landmark matching is shown in Fig. 4.10. Traditional landmark matching is shown on the left side (green), while GDAS landmark matching is shown in the center column (blue). In the right hand column, the surfaces are showed overlaid on a T1 image, with the gold standard segmentation shown in red.

Qualitatively, the improvement of the GDAS algorithm over traditional landmark matching is evident. Large distortions at the head of the hippocampus are common where landmark placement can be quite variable. Along the length of the hippocampus, deformations with scale characteristic of the distance between landmarking planes are easily seen. These issues are still common in those surfaces not requiring quality control. The GDAS algorithm avoids each of these pitfalls, avoiding overfitting landmarks while maintaining shape variability characteristic of the population.

Evaluation: ROI method case study

For the landmark matching application, we describe in detail the performance of the GDAS algorithm as compared to our existing method.

The overlap on a large scale is quantified by κ scores, as shown in Fig. 4.11 for each of the 16 test cases. The GDAS results tend to be similar, but better on average than that for landmark LDDMM. For typical landmark matching the mean and standard deviation of κ is 0.7131 ± 0.0457, and for GDAS landmark matching it is 0.7268 ± 0.0531. The difference is statistically significant ($p < 0.05$ in paired t-test).

For those cases with two raters, we examine the second κ score, which differed from the first by 2.66% on average, to understand inter-rater variability. We present κ scores, averaged over the two raters in Table 4.1. In each case, the GDAS method performs superiorly for both raters, and this is reflected the increased average κ scores from 0.732 to 0.751. Despite this improvement, it is interesting to note that the κ overlap between the two manual segmentations is comparable to that between the results of the two segmentation methods.

Examining overlap voxel by voxel, as in Fig. 4.11, shows our algorithm making a small improvement in accuracy. However, the relatively larger improvement in robustness can be seen when examining surface shapes globally such as in Fig. 4.10, and contrasting with expectations from a knowledge of neuroanatomy. The region around the head was seen to be particularly challenging to

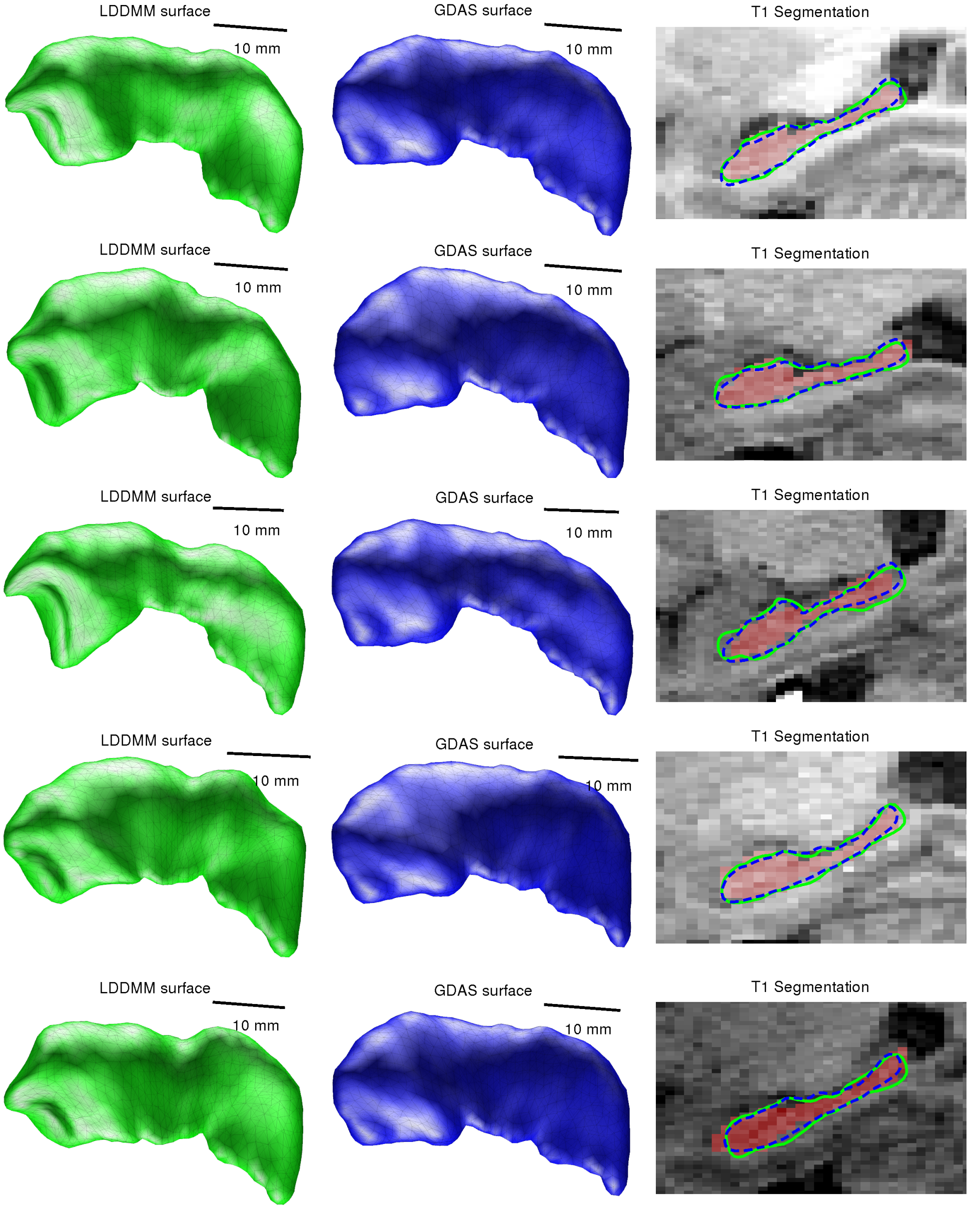


Figure 4.10: Segmentation results for standard landmark matching (left/green/solid), and GDAS landmark matching (center/blue/broken) for 5 examples (first 3 were identified for quality control, final 2 were not). Segmentations overlaid with corresponding T1 image and "ground truth" (red highlight) are shown in the right column.

Table 4.1: Inter-rater variability is examined by presenting κ overlap between various pairs of data (indicated in the left column).

Case	1	2	3	4	5	Average
LDDMM vs. Manual Segs.	0.780	0.743	0.718	0.719	0.702	0.732
GDAS vs. Manual Segs.	0.784	0.768	0.752	0.721	0.731	0.751
Manual Segs.	0.866	0.853	0.813	0.800	0.784	0.823
LDDMM vs. GDAS	0.832	0.820	0.830	0.827	0.829	0.828

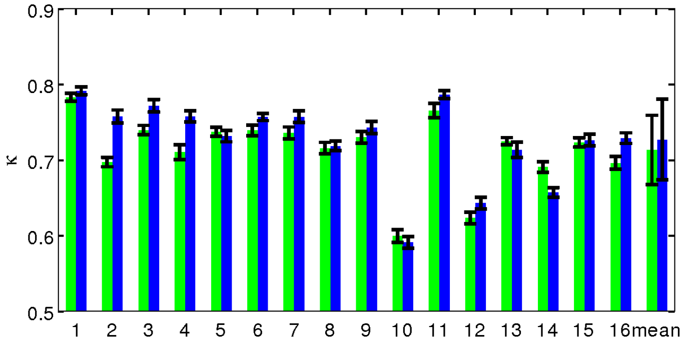


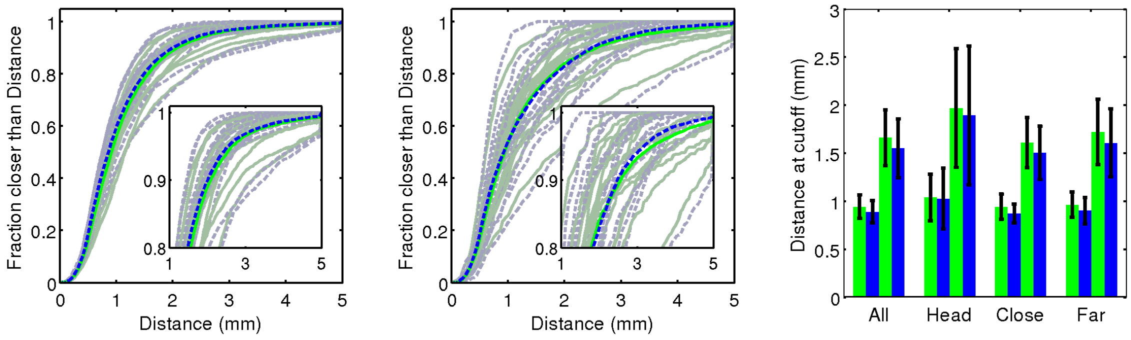
Figure 4.11: Kappa scores are shown for each of the 16 patients examined, with mean and standard deviation shown on the right. Green/left of pair: typical landmark matching, blue/right of pair: GDAS landmark matching.

segment in the traditional landmark case, and distortions occurring at the scale of landmark spacing give the impression that certain regions are "left behind". To quantify accuracy globally, while acknowledging these specifically challenging areas, we use surface-to-surface-distance histograms and associated CDFs.

These CDFs are shown for all 16 patients (unsaturated colors, green: standard landmark matching, blue: GDAS) in Fig. 4.12a. Combining all vertices gives a single CDF indicative of the whole population (saturated colors). A CDF closer to the top left reflects a better segmentation. In Fig. 4.12b we show the same analysis, but restricted to vertices within 10 mm of the head landmark. This analysis was repeated (not shown) with vertices restricted to those within 2.5 mm of any landmark, and those not within 2.5 mm of any landmark.

For each patient, the 50% and 80% crossings were measured, and are plotted in Fig. 4.12c. In each set of four bars, the left two show 50% crossings, and the right two show 80% crossings. A smaller value indicates a better segmentation, but the 50% crossing indicates a "typical" region, while the 80% crossing indicates an "outlier" region. Our hypothesis was that the GDAS algorithm would show improvement in outlier regions, at the cost of poorer performance in typical regions. However, the data shows better performance from GDAS in all regions examined. This is likely due to traditional LDDMM overfitting landmark placement inaccuracies, while GDAS finds an appropriate balance between landmark matching accuracy and shape variability. Differences show statistical significance ($p < 0.05$ in a paired t-test) with the exception of vertices close to the head (50%: $p = 0.4073$, 80%: $p = 0.0895$).

To further quantify the more natural shapes produced by GDAS, we examine the curvature of the resulting segmentations. For each patient examined, the integrated sum of squares of principal curvatures is shown in Fig. 4.13. In all but one case, the GDAS algorithm results in surfaces with less curvature. The differences are statistically significant ($p < 0.0001$ in a paired t-test).



(a)

(b)

(c)

Figure 4.12: Surface to surface distance CDFs including (a) all vertices, and (b) only vertices within 10 mm of head landmark. Inset shows same data zoomed to ≥80%. Plot (c) shows the 50% (left pair in a set of four) and 80% (right pair in a set) crossing for the vertices shown in (a) ("All") and (b) ("Head"), as well as within 2.5 mm of any landmark ("Close") or not ("Far"). Green/solid/left of pair: traditional LDDMM, blue/broken/right of pair: GDAS.

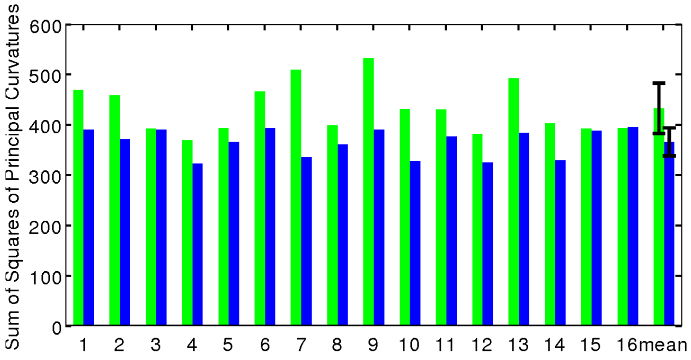


Figure 4.13: Integrated sum of squares of principal curvatures is shown for the 16 patients examined, as well as means and standard deviations. Green/left of pair: typical landmark matching, blue/right of pair: GDAS landmark matching.

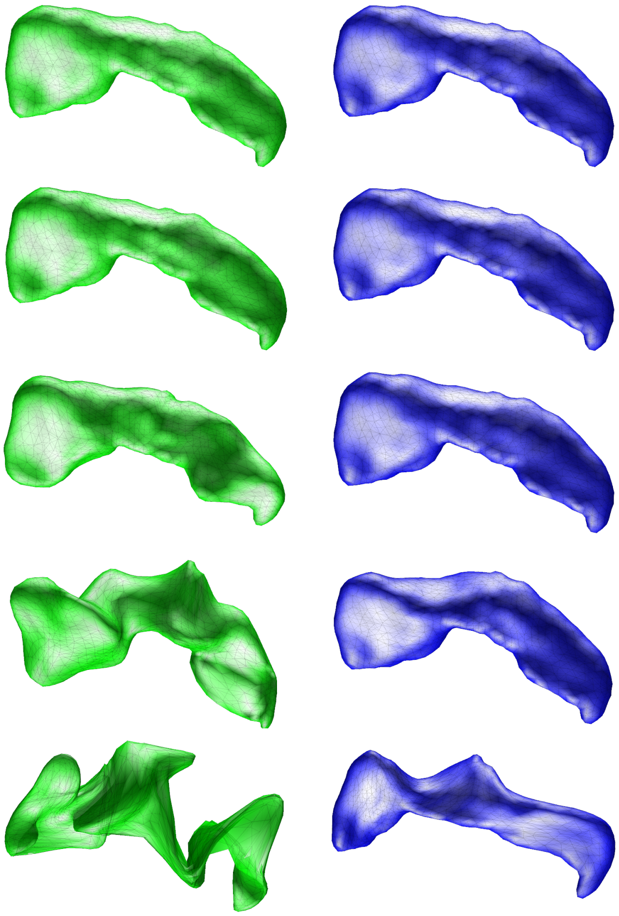


Figure 4.14: Example result from simulated data. Left: Standard landmark matching, Right: GDAS landmark matching. Landmark variance from top to bottom: 0.004475, 0.04475, 0.4475, 4.475, 44.75.

Evaluation: Simulated data

To further quantify the performance and robustness of our landmark matching algorithm, we evaluate it using simulated data such that the gold standard segmentation can be precisely known. Figure 4.14 shows example results of our landmark matching algorithms as described above, with traditional landmark matching shown on the left side (green), and GDAS landmark matching shown on the right side (blue). From top to bottom, the additive noise in landmark placement increases from 1/10 to 10 times that expected from voxel size in our case study (variance 0.004475, 0.04475, 0.4475, 4.475, 44.75). At low levels of landmark uncertainty, the two algorithms give very similar results.

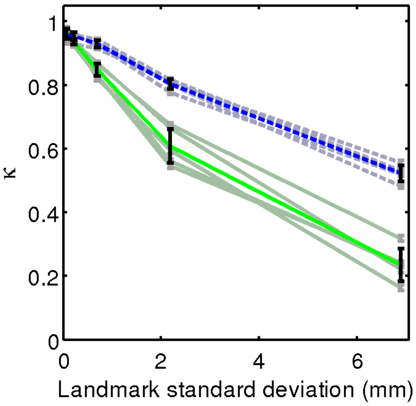


Figure 4.15: Kappa scores are shown for the 5 simulated cases (desaturated), and mean and standard deviation (saturated), as a function of landmark noise. Green/bottom: standard landmark matching, blue/top: GDAS landmark matching.

However, as landmark uncertainty increases, the performance of GDAS exhibits a graceful decline, while that of traditional LDDMM demonstrates a precipitous drop. Note that third row gives a level of landmark uncertainty comparable to that in our case study.

Figure 4.15 shows κ scores as a function of landmark noise for each of the five simulated cases (desaturated colors), as well as for the average performance (saturated colors). Consistent with our expectations of improved robustness, we see a much smaller variability in κ scores for GDAS. Furthermore, consistent with our earlier discussion of accuracy, we see poor performance of traditional LDDMM due to overfitting untrustworthy data. Note that the third datapoint (close to the left hand side of the figure) corresponds to a level of landmark uncertainty comparable to that in our case study.

For the GDAS results, we can also express accuracy by measuring the error in PCA coefficients recovered by the algorithm. A natural way to do this is through the Mahalanobis distance (treating the inverse of the covariance matrix as an bilinear symmetric operator defining an inner product). Loosely, this distance is the square root of the sum of squares of differences in PCA coefficients, each divided by its respective standard deviation. At the five levels of landmark noise examined, the distance between the true coefficients and those recovered by GDAS (summed over the 31 coefficients), are given by 0.8100, 2.0484, 4.0477, 5.5894, 5.7708 standard deviations. However, the lower order coefficients, which contribute more to overall shape, are recovered with more accuracy than the higher ones. The first coefficient is recovered with an error of 0.0154, 0.0259, 0.1422, 0.4338, 0.6215 standard deviations, and the first 5 with an error of 0.0813, 0.2791, 0.5213, 1.6052, 2.0663 standard deviations.

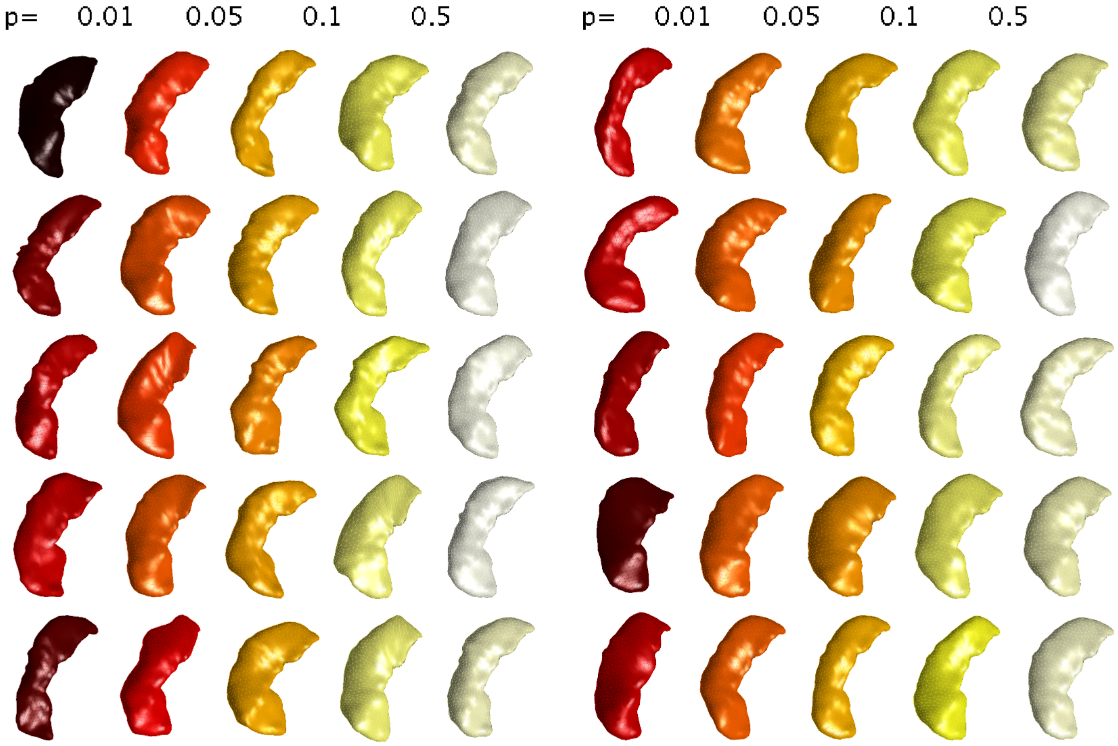
This highlights a potential future direction for the GDAS framework. We calculate the Mahalanobis distance from the origin for each of the 650 patients in our ADNI training set, and use the empirical distribution to calculate p-values. A sample of these patients is shown in Fig. 4.16a. Surfaces are colored by their p-value, and binned for p between the values $\{0, 0.01, 0.05, 0.1, 0.5, 1\}$. Each column represents one bin, and five example cases per bin are shown. It is evident that such a distance is descriptive of the naturalness of anatomical shapes, with low p-values corresponding to unnatural shapes. Using such a tool to identify outliers for targeted quality control is the subject of future research.

An extension of this idea is the ability to generate random anatomical shapes, and quantify their typicality with p-values. Some examples, are shown in Fig. 4.16b, with format paralleling that discussed above. This tool demonstrates the generative nature of the GDAS framework, and may prove to be useful for didactic or other purposes.

4.6.3 Inside-outside modeling

We seek to automatically segment subcortical structures from MR images. For simplicity, we assume that such structures are relatively homogeneous throughout, and therefore chose an appearance model for voxel intensities that depend on location only through whether they are inside the structure or not. To perform image segmentation, we seek to partition the space into high integrated voxel-likelihood under such an inside-outside model. The approach can be generalized to more complex appearance models (involving higher order image features, for example) in a straightforward way.

Gaussian mixtures are widely employed to model voxel intensity of medical imaging. That is, a



(a) (b)

Figure 4.16: Population of left hippocampus surfaces binned by p-value from ADNI dataset. a) Real patient data. b) Simulated data.

p.d.f of intensity I at a certain location in the tissue is in the form of

$$p(I; x) = \sum_{q=1}^{Q} \pi_q(x) \frac{1}{\sqrt{2\pi\sigma_q^2(x)}} \exp\left(-\frac{(I - \mu_q(x))^2}{2\sigma_q^2(x)}\right)$$

where π_q, μ_q, σ_q^2 denote the weight, mean and variance of q-th (out of Q) Gaussian component respectively.

In our work, we assume that the intensities at all points of the interior region (resp. exterior region) of the surfaces share the same mixed Gaussian distribution and the p.d.f's are denoted as p_{int} and p_{ext} respectively.

Given the number of mixture components, the maximum likelihood estimator for the parameters can be computed using the EM algorithm [42]. Our estimation of p_{int} and p_{ext} (using mixtures of Gaussians) is performed on the basis of training images with manual segmentation, in which the collection of all intensity values of voxels inside (resp. outside) the ROI are used for p_{int} (resp. p_{ext}).

We define the accuracy of segmentation using integrals of likelihood of being misclassified, and we define the mismatch:

$$\begin{aligned} E_2 = &\int_{x \in int(S)} \log(p_{ext}(I(x)))dx \\ &+ \int_{x \in ext(S)} \log(p_{int}(I(x)))dx \end{aligned} \tag{4.2}$$

where we denote the interior and exterior of a closed surface S by $int(S)$ and $ext(S)$ respectively. The constant σ^2 in (4.2) is determined heuristically. The gradient of this function depends only on f_1 and is derived in [25].

Methods

To demonstrate the capabilities of the GDAS image segmentation algorithm, 5 examples for the same dataset as the landmark matching study are shown. We anticipate that good initial alignment will be important for high quality segmentations, and so the same landmark based similitude registration as above will be used to initialize the target data in this study.

We use 4 outside and 3 inside components for our Gaussian mixture model. The mixture model is trained based on gold standard segmentations from the remaining cases in a “leave-one-out” fashion.

A histogram equalization intensity transformation is applied to each T1 image to match the first training sample, based on data from a neighborhood (± 5 voxels) around the landmarks, before estimating mixture model coefficients. A similar histogram equalization is applied to the target image (to match the first training sample) before beginning the segmentation process.

Results

An example of the results of Gaussian mixture modeling are shown as probability density functions in Fig 4.18. Measured data (after histogram equalization) is shown as a solid curve, and the results of mixture modeling as dashed curve. The Gaussian mixture parameters are quite similar in all cases examined. The "inside" region (narrow curve, blue and red) is a unimodal distribution describing subcortical gray matter. The challenge of this application can be seen from the "outside" region (broad curve, green and magenta), which is a more heterogeneous mixture. It describes cerebrospinal fluid and white matter, as well as cortical gray matter and partial-volume voxels whose intensities are quite similar to "inside".

Five example segmentation results are shown in Fig. 4.17. The performance appears satisfactory, an achievement considering the large overlap between inside and outside histograms seen in Fig. 4.18. The PCA prior can prevent the template surface from deforming to erroneously include cortical gray matter in many cases, even though it is similar or identical in intensity to subcortical gray matter. This simple inside/outside model could likely be improved. For example by including a heterogeneous appearance model, or combining landmarks and intensity information in cost functions. However, this will be the subject of future research. The purpose of this section was to demonstrate the extensibility of the GDAS framework to a varied range of applications.

4.7 Discussion

Volumetric segmentation has played an essential role in computer-based interpretation of medical images. There have been many approaches published to address this challenge [16, 43–50]. Most segmentation methods use a combination of shape constraint and data attachment to achieve their

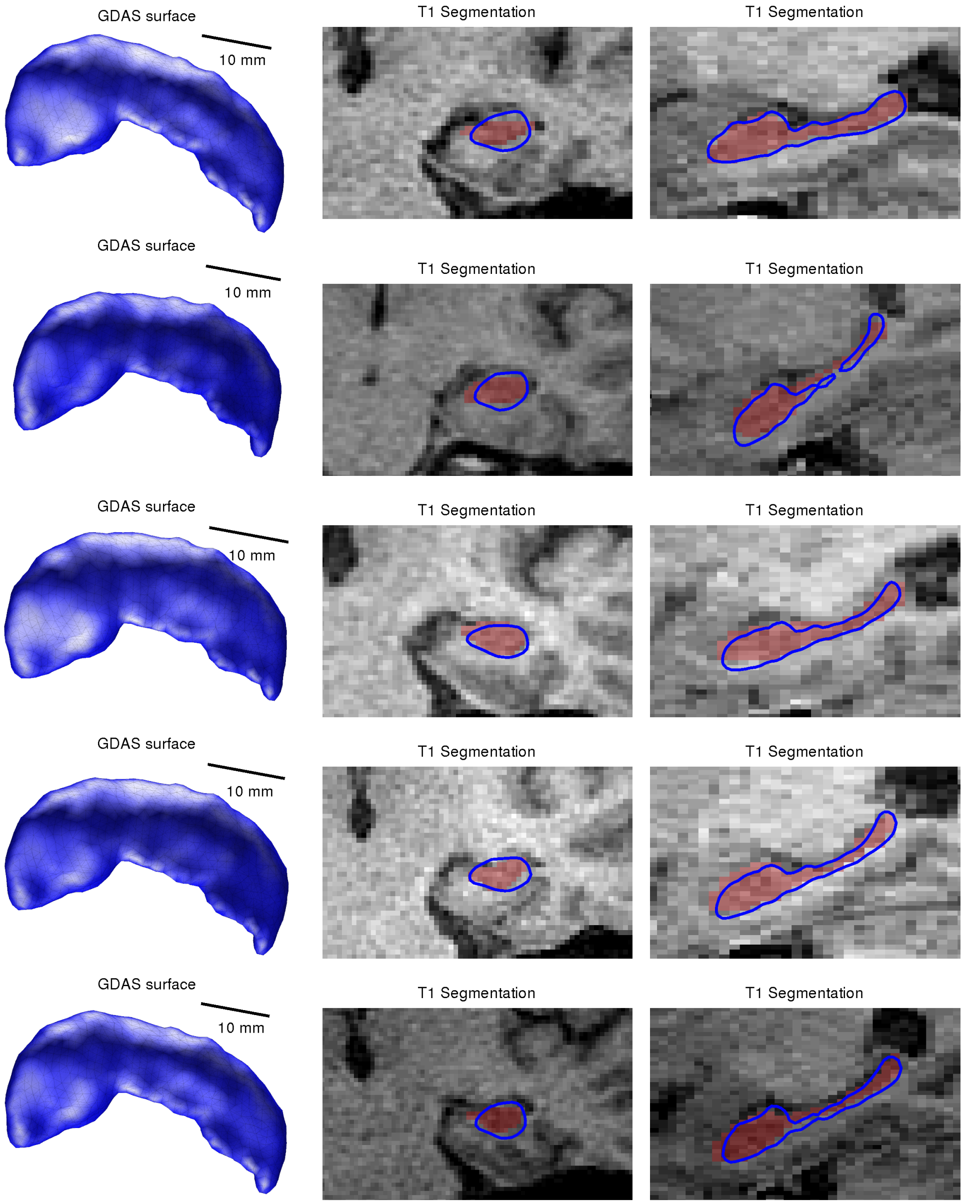


Figure 4.17: Example segmentations of T1 images using GDAS image segmentation based on inside-outside modeling. The resulting surfaces are shown on the left hand side, and T1 images with gold standard (red highlight) and segmentation (blue curve) are shown in coronal (center) and sagittal (right) views.

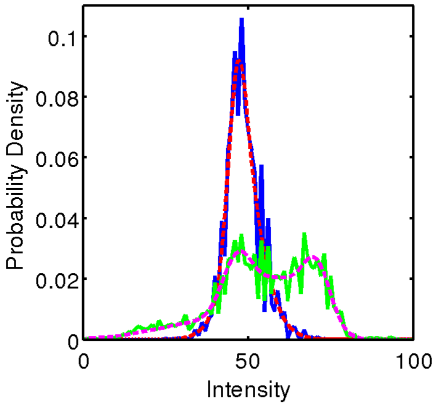


Figure 4.18: Mixture modeling is shown for inside (narrow curve, red/blue), and outside (broad curve, magenta/green) T1 voxel intensities (after histogram equalization). The T1 data is shown with broken lines, and the mixture model with solid lines.

goals. Data attachment can be based on geometric data, gray levels, edge detection [51], or unstructured segmentation like K-Means or Gaussian Mixtures[52].

In this paper, we have demonstrated applications of the GDAS framework with data attachment based on landmarks, surfaces, and likelihood ratios from grayscale values. For the case of landmark matching, we demonstrated how it could be used to remove or hasten a laborious quality control phase of large scale neuroanatomical studies. We quantified its improvement over an existing method based on accuracy, as well as robustness. As is typical of Bayesian analysis, we originally expected to see a trade off between accuracy and robustness. However, our results showed improvements in both, likely due to overfitting to noisy data in the standard method, reducing accuracy.

The GDAS algorithm shows improvement over traditional landmark matching in κ scores and surface to surface distances, as compared to the gold standard segmentation. Qualitatively, improvements are particularly noticeable in the region around the hippocampus' head (where landmark placement is uncertain). The segmentations resulting from GDAS appear natural, reflecting the typicality and variability of the population from which the PCA basis was determined. This naturalness was quantified in terms of reduced curvature as compared to the traditional method, and can be understood in terms of Mahalanobis distance p-values.

We have found in large sample studies that robustness is accommodated by our GDAS methods controlled by the PCA dimensions empirically trained from samples of subcortical anatomy. In a study with over 400 hand placements of landmarks in hippocampus and amygdala, we have found that robust GDAS detects our failed landmark based mappings using p-values, supporting the notion that it provides direct method for quality control of large deformation mappings. Exploring this possibility will be the subject of future research.

Our method is based on the geodesically controlled diffeomorphism constraints associated with the momentum conservation law. Encoding structure via prior distributions which are empirically trained has a longstanding tradition in active shape and appearance modeling [31, 53], defined on landmark structures as well as on higher dimensional structures as proposed in [54–56]. Our principal contribution here has been to encode the diffeomorphism constraint into the standard active shape models. By incorporating the conservation law controls, we not only inherit the power of diffeomorphic transfer of the submanifold surface in the background 3D space, as has been described in [57, 58], but we also obtain the metric structure property. Along the geodesic path connecting templates and targets, the metric structure of the large space is maintained.

These properties have been explicitly modeled in our own methods previously using deformable templates acted upon by diffeomorphisms and embedding them into the associated metric space structures [19, 59]. These formulations have tended to explicitly model the transformations on the entire dense background space Ω, working to minimize a cost function accumulated over the entire space. In the setting where the template consists of a collection of homogeneous substructures, we would expect to obtain similar formulations as described herein. In fact, Qiu et al.[34] have used dense deformations for statistical modeling in such settings via their support on the boundaries of the cortical substructures.

Chapter 5

Image matching with geodesic parameters

While surface matching with geodesic parameters proved successful in several applications, the algorithm is limited in that the data fidelity function must depend only on the location of surface vertices. For matching images using this representation of shape, another approach must be devised. Large numbers of datasets are becoming available which are in the form of dense binary segmentations, which represent substructures with smooth surface boundaries. Methods which can include the dense segmentation layers in the mapping will be invaluable.

5.1 Cost function and dynamical constraints

Here we describe the state of our flow with three variables. The singular geodesic coordinates p, the surface parameterization $f : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ with $f_t = \varphi_t(f_0)$, and an image I which is a function from $\Omega \subset \mathbb{R}^3 \to \mathbb{R}^C$ (i.e. a set of C channels in an image are modeled as a vector valued image).

We have dynamics

$$\frac{d}{dt}\begin{pmatrix} f \\ I \\ p \end{pmatrix} \doteq \Lambda(f, I, p) = \begin{pmatrix} v(f) \\ -DIv \\ -Dv^T(f)p \end{pmatrix} = \begin{pmatrix} \int_U K(f(u), f(u'))p(u')du' \\ -DI \int_U K(f(u), f(u'))p(u')du' \\ -\int_U p^T(u)p(u')\nabla_1 K(f(u), f(u'))du' \end{pmatrix}$$

where the image dynamics are given by optical flow as shown in C.3. As before we formulate the matching as a constrained optimization problem.

Problem 4 (Image matching with singular geodesic coordinates)**.** *Given a template surface parameterization f_0 and template image I_0, some data fidelity cost $F(f_1, I_1)$ (e.g. sum of square error between I_1 and a target), and some regularization cost R (e.g. norm squared of v_0), we seek to minimize the cost*

$$C(p_0) = R(p_0) + F(f_1, I_1)$$

over the singular geodesic coordinates p_0 subject to the dynamical constraints $\frac{d}{dt}f = K_{ff}p = v(f)$, $\frac{d}{dt}I = -DIv$ and $\frac{d}{dt}p = -Dv^T(f)p$.

As before we introduce Lagrange multipliers to form the augmented cost function

$$\begin{aligned} &R(p_0) + F(f_1, I_1) \\ &\quad + \int_0^1 (\lambda^f | \frac{d}{dt}f - \int_U K(f(u), f(u'))p(u')du') \\ &\quad + (\lambda^I | \frac{d}{dt}I + DI(x)\int_U K(x, f(u'))p(u')du') \\ &\quad (\lambda^p | \frac{d}{dt}p + \int_U p^T(u)p(u')\nabla_1 K(x, f(u'))du')dt \end{aligned} \tag{5.1}$$

5.2 Necessary conditions

Statement 4 (Necessary conditions for image matching with singular geodesic coordinates)**.** The necessary conditions for stationary solutions to problem 4 satisfy the following dynamical equations:

$$\begin{aligned} \frac{d}{dt}\lambda^f(u) = &\int_U [\lambda^{fT}(u)p(u') + \lambda^{fT}(u')p(u)]\nabla_1 K(f(u), f(u'))du' \\ &+ \int_\Omega \nabla_1 K(f(u), x)p^T(u)DI^T(x)\lambda^I(x)dx \end{aligned}$$

$$+ \int p^T(u')p(u)[D_1\nabla_1 K(f(u), f(u'))\lambda_p(u) + D_2\nabla_1 K(f(u), f(u'))\lambda_p(u')]du' \tag{5.2}$$

$$\frac{d}{dt}\lambda^I(x) = -\operatorname{div}[v(x)\lambda^T(x)]^T \tag{5.3}$$

$$\begin{aligned}\frac{d}{dt}\lambda^p(u) = &-\int_U K(f(u), f(u'))\lambda^f(u')du' + \int_\Omega K(f(u), x)DI^T(x)\lambda^I(x)dxdu' \\ &+ \int p(u')[D_2K(f(u), f(u'))\lambda^p(u') + D_1K(f(u), f(u'))\lambda^p(u)]du'\end{aligned} \tag{5.4}$$

or for stationary kernels

$$\begin{aligned}\frac{d}{dt}\lambda^f(u) = &-\int_U [\lambda^{fT}(u)p(u') + \lambda^{fT}(u')p(u)]\nabla_1 K(f(u), f(u'))du' \\ &+ \int_\Omega \nabla_1 K(f(u), x)p^T(u)DI^T(x)\lambda^I(x)dx \\ &+ \int p^T(u')p(u)D_1\nabla_1 K(f(u), f(u'))[\lambda^p(u) - \lambda^p(u')]du' \\ \frac{d}{dt}\lambda^I(x) = &-\operatorname{div}[v(x)\lambda^T(x)]^T \\ \frac{d}{dt}\lambda^p(u) = &-\int_U K(f(u), f(u'))\lambda^f(u')du' \\ &+ \int_\Omega K(f(u), x)DI^T(x)\lambda^I(x)dxdu' \\ &+ \int p(u')D_1K(f(u), f(u'))[\lambda^p(u) - \lambda^p(u')]du'\end{aligned}$$

which is simplified by the explicit solution for λ_t^I

$$\lambda_1^I = \lambda_1^I(\varphi_{1t}^{-1})|D\varphi_{1t}^{-1}|$$

and the boundary conditions $\lambda_1^f = -D_1F^T(f_1, I_1)$, $\lambda_1^I = -D_2F^T(f_1, I_1)$, $\lambda_1^p = 0$, $\lambda_0^p = DR^T(p_0)$.

As compared to surface matching only (3.4) and (3.5), the dynamics for λ^f and λ^p each have an extra term added. These extra terms allow them to “feel” the deforming image in a volume determined by the kernel K.

Proof. The augmented cost function (5.1) must be stationary to perturbations in each variable. In what follows, there will be a subsection for perturbations for each variable, and sub sub sections describing each term in detail.

5.2.1 Perturbation with respect to f

We perturb the augmented cost function of (5.1) with respect to f, removing terms that do not vary with the perturbation for compactness.

$$\begin{aligned}
&\frac{d}{d\epsilon}F(f_1+\epsilon\delta f_1, I_1) + \int_0^1 (\lambda^f|\frac{d}{dt}f + \epsilon\frac{d}{dt}\delta f) \\
&\quad -(\lambda^f|\int_U K(f(u)+\epsilon\delta f(u), f(u')+\epsilon\delta f(u'))p(u')du') \\
&\quad -(\lambda^I| - DI(x)\int_U K(x, f(u')+\epsilon\delta f(u'))p(u')du') \\
&\quad -(\lambda^p| - \int_U p^T(u)p(u')\nabla_1 K(f(u)+\epsilon\delta f(u), f(u')+\epsilon\delta f(u'))du')dt\Bigg|_{\epsilon=0}
\end{aligned}$$

Perturbation of boundary terms and $\frac{d}{dt}f$ with respect to f

Examining these terms gives

$$\begin{aligned}
&(D_1F^T(f_1,I_1)|\delta f_1) + \int(\lambda^f|\frac{d}{dt}\delta f)dt \\
&\quad = (D_1F^T(f_1,I_1)|\delta f_1) - \int(\frac{d}{dt}\lambda^f|\delta f)dt + (\lambda_1^f|\delta f_1)
\end{aligned}$$

after applying integration by parts and recalling $\delta f_0 = 0$.

For a stationary solution, the boundary term implies that $\lambda_1^f = -D_1F^T(f_1, I_1)$. The term $\frac{d}{dt}\lambda^f$ will be balanced against those below to achieve stationarity.

Perturbation of f dynamics with respect to f

We suppress the integral over time here.

$$\begin{aligned}
&\frac{d}{d\epsilon} - (\lambda^f|\int_U K(f(u)+\epsilon\delta f(u), f(u')+\epsilon\delta f(u'))p(u')du')\Bigg|_{\epsilon=0} \\
&\quad = (-\int_U [\lambda_f^T(u)p(u') + \lambda_f^T(u')p(u)]D_1K(f(u),f(u'))^T du'|\delta f(u))
\end{aligned}$$

This term was calculated in section 3.2.1.

Perturbation of I dynamics with respect to f

$$\frac{d}{d\epsilon}(\lambda^I|DI(x)\int_U K(x, f(u')+\epsilon\delta f(u'))p(u')du')\Bigg|_{\epsilon=0}$$

$$
\begin{aligned}
&= (\lambda^I | DI(x) \int_U D_2 K(x, f(u'))\delta f(u') p(u') du') \\
&= \int_\Omega \int_U \lambda^{IT}(x) DI(x) D_2 K(x, f(u'))\delta f(u') p(u') du' dx \\
&= (\int_\Omega \nabla_2 K(x, f(u')) p^T(u') DI^T(x) \lambda^I(x) dx | \delta f(u')) \\
&= (\int_\Omega \nabla_1 K(f(u'), x) p^T(u') DI^T(x) \lambda^I(x) dx | \delta f(u'))
\end{aligned}
$$

Perturbation of p with respect to f

$$
\begin{aligned}
&\frac{d}{d\epsilon}(\lambda^p | \int_U p^T(u) p(u') \nabla_1 K(f(u) + \epsilon\delta f(u), f(u') + \epsilon \delta f(u')) du')\Big|_{\epsilon=0} \\
&\quad = (\int_U p^T(u') p(u) [D_1 \nabla_1 K(f(u), f(u'))\lambda_p(u) + D_2 \nabla_1 K(f(u), f(u'))\lambda_p(u')] du' | \delta f(u))
\end{aligned}
$$

This term was calculated in 3.2.1. For stationary kernels it becomes

$$
(\int_U p^T(u') p(u) D_1 \nabla_1 K(f(u), f(u')) [\lambda_p(u) - \lambda_p(u')] du' | \delta f(u))
$$

Determining λ^f dynamics

Putting this together gives

$$
\begin{aligned}
\frac{d}{dt}\lambda^f = &\int_U [\lambda^{fT}(u) p(u') + \lambda^{fT}(u') p(u)] \nabla_1 K(f(u), f(u')) du' \\
&+ \int_\Omega \nabla_1 K(f(u), x) p^T(u) DI^T(x) \lambda^I(x) dx \\
&+ \int_U p^T(u') p(u) [D_1 \nabla_1 K(f(u), f(u'))\lambda_p(u) + D_2 \nabla_1 K(f(u), f(u'))\lambda_p(u')] du'
\end{aligned}
$$

Or for stationary kernels

$$
\begin{aligned}
\frac{d}{dt}\lambda^f = &-\int_U [\lambda^{fT}(u) p(u') + \lambda^{fT}(u') p(u)] \nabla_1 K(f(u), f(u')) du' \\
&+ \int_\Omega \nabla_1 K(f(u), x) p^T(u) DI^T(x) \lambda^I(x) dx \\
&+ \int_U p^T(u') p(u) D_1 \nabla_1 K(f(u), f(u')) [\lambda^p(u) - \lambda^p(u')] du'
\end{aligned}
$$

5.2.2 Perturbation with respect to I

We perturb the augmented cost function (5.1) with respect to I, removing terms that do not vary with the perturbation for compactness.

$$\frac{d}{d\epsilon}F(f_1, I+\epsilon\delta I_1) + \int_0^1 (\lambda^I|\frac{d}{dt}I) \\ - (\lambda^I| - (DI(x)+\epsilon D\delta I(x))v(x))dt\bigg|_{\epsilon=0}$$

Perturbation of boundary terms and $\frac{d}{dt}I$ with respect to I

$$(D_2F^T(f_1,I_1)|\delta I_1) - \int(\frac{d}{dt}\lambda^I|\delta I)dt + (\lambda_1^I|\delta I_1)$$

after applying integration by parts and recalling $\delta I_0 = 0$. This gives the boundary condition

$$\lambda_1^I = -D_2F^T(f_1, I_1)$$

Perturbation of f dynamics with respect to I

This term is 0.

Perturbation of I dynamics with respect to I

$$(\lambda^I|D\delta Iv) = (-\operatorname{div}[\lambda^{IT}v]^T|\delta I)$$

Perturbation of p dynamics with respect to I

This is 0.

Determining λ^I dynamics

Putting this together gives

$$\frac{d}{dt}\lambda^I = -\operatorname{div}[v(x)\lambda^I(x)]$$

Note that as shown in Sec. C.4 this equation has an explicit solution in terms of the diffeomorphisms, $\lambda_t^I = \lambda_1^I(\varphi_t^{-1})|D\varphi_t^{-1}|$.

In our problem we will deform from $t = 1$, so we will use

$$\lambda_t^I = \lambda_1^I(\varphi_{1t}^{-1})|D\varphi_{1t}^{-1}|$$

where $\varphi_{1t} = \varphi_t(\varphi_1^{-1})$ (i.e. from time 1 back to time zero, and then from time 0 to time t).

5.2.3 Perturbation with respect to p

We perturb the augmented cost (5.1) with respect to p, removing terms that do not vary with the perturbation.

$$\begin{aligned}\frac{d}{d\epsilon} R(p_0 + \epsilon\delta p_0) + \int (\lambda^p | \frac{d}{dt} p + \epsilon \frac{d}{dt}\delta p)\\ - (\lambda^f | \int_U K(f(u), f(u'))(p(u')) + \epsilon\delta p(u')du')\\ - (\lambda^I | - DI(x)\int_U K(x, f(u'))(p(u') + \epsilon\delta p(u'))du')\\ - (\lambda^p | - \int_U (p(u) + \epsilon\delta p(u))^T (p(u') + \epsilon\delta p(u'))\nabla_1 K(x, f(u'))du' dt \Big|_{\epsilon=0}\end{aligned}$$

Perturbation of boundary terms and $\frac{d}{dt}p$ with respect to p

Examining these terms gives

$$\begin{aligned}(DR^T(p_0)|\delta p_0) + \int (\lambda^p | \frac{d}{dt}\delta p)dt\\ = (DR^T(p_0)|\delta p_0) - \int (\frac{d}{dt}\lambda^p|\delta p)dt + (\lambda_1^p|\delta p_1) - (\lambda_0^p|\delta p_0)\end{aligned}$$

after applying integration by parts. For stationarity with respect to perturbations, this implies the boundary conditions $\lambda_1^p = 0$ and $\lambda_0^p = DR^T(p_0)$.

The term $\frac{d}{dt}\lambda^p$ will be balanced against those below to achieve stationarity.

Perturbation of f dynamics with respect to p

$$\begin{aligned}-(\lambda^f | \int_U K(f(u), f(u'))\delta p(u')du')\\ = (-\int K(f(u), f(u'))\lambda^f(u')du' | \delta p(u))\end{aligned}$$

This term was calculated above in section 3.2.2.

Perturbation of I dynamics with respect to p

$$\begin{aligned}
&(\lambda^I|DI(x)\int_U K(x,f(u'))\delta p(u')du') \\
&\quad= \int_\Omega \lambda^{IT}(x)DI(x)\int_U K(x,f(u'))\delta p(u')du'dx \\
&\quad= \int_U\int_\Omega \lambda^{IT}(x)DI(x)K(x,f(u'))dx\delta p(u')du' \\
&\quad= (\int_\Omega K(f(u'),x)DI^T(x)\lambda^I(x)dx|\delta p(u'))
\end{aligned}$$

Perturbation of p dynamics with respect to p

$$\begin{aligned}
&(\lambda^p|\int_U(\delta p^T(u)p(u')+\delta p^T(u')p(u))\nabla_1 K(x,f(u'))du') \\
&\quad= (\int p(u')[D_2K(f(u),f(u'))\lambda^p(u')+D_1K(f(u),f(u'))\lambda^p(u)]du'|\delta p(u))
\end{aligned}$$

This was calculated above in section 3.2.2.

For stationary kernels this gives

$$(\int p(u')D_1K(f(u),f(u'))[\lambda_p(u)-\lambda_p(u')]du'|\delta p(u))$$

Determining λ^p dynamics

Putting this together gives

$$\begin{aligned}
\frac{d}{dt}\lambda^p = &-\int_U K(f(u),f(u'))\lambda^f(u')du' \\
&+\int_\Omega K(f(u),x)DI^T(x)\lambda^I(x)dxdu' \\
&+\int p(u')[D_2K(f(u),f(u'))\lambda^p(u')+D_1K(f(u),f(u'))\lambda^p(u)]du'
\end{aligned}$$

or for stationary kernels

$$\begin{aligned}
\frac{d}{dt}\lambda^p = &-\int_U K(f(u),f(u'))\lambda^f(u')du' \\
&+\int_\Omega K(f(u),x)DI^T(x)\lambda^I(x)dxdu' \\
&+\int p(u')D_1K(f(u),f(u'))[\lambda^p(u)-\lambda^p(u')]du'
\end{aligned}$$

5.3 Gradient of the cost function

As before, with all constraints obeyed we have $\frac{d}{dp_0}F(f_1, I_1) = -\lambda_0^p$, so that $\frac{d}{dp_0}C = DR^T(p_0) - \lambda_0^p$. We can write this as a gradient vector

$$\begin{aligned}(DR^T(p_0) - \lambda_{p_0}|\delta p_0) &= (K_{ff}K_{ff}^{-1}(DR^T(p_0) - \lambda_{p_0}))|\delta p_0) \\ &= \langle K_{ff}^{-1}(DR^T(p_0) - \lambda_{p_0}), \delta p_0\rangle_{V^*}\end{aligned}$$

Where we are identifying functions on U with singular distributions in V^*.

Typically $R(p_0) = \frac{1}{2}\iint K(f_0(u), f_0(u'))p^T(u)p(u')du du'$. In this case $DR(p_0) = \int K(f_0(u), f_0(u'))p_0^T(u')du'$ and so the gradient vector is just p_0. Typically $F(f_1, I_1) = \frac{1}{2}\int |I_1(x) - J(x)|^2 dx$ for J a target image, so that $D_1F(f_1, I_1) = 0$.

5.4 Algorithm for image matching with singular geodesic coordinates

This gives the following gradient descent based algorithm for image matching

Algorithm 5.1. Image matching with singular geodesic coordinates

```
1. Initialize $f_0$ to a template, initialize $p_0$ (typically to 0)
2. Calculate $f_1, p_1$ by flowing $\frac{d}{dt}f = v(f)$ and $\frac{d}{dt}p = -Dv^T(f)p$ forward in time
3. Calculate $v_t(x)$, $\varphi_t^{-1}(x)$ and $\varphi_{1t}^{-1}(x)$ sampled on a dense grid
4. Calculate $DF(f_1, I_1)$
5. Initialize boundary conditions $\lambda_1^f = -D_1F^T(f_1, I_1)$, $\lambda_1^p = 0$, $\lambda_1^I = -D_2F(f_1, I_1)^T$
6. Calculate $\lambda_t^I$ explicitly as $\lambda_1^I(\varphi_{1t}^{-1})|\varphi_{1t}^{-1}|$
7. Flow $\lambda_f, \lambda_p$ backwards to $t = 0$ using (5.2) and (5.4)
8. Calculate gradient vector $K_{ff}^{-1}(DR(p_0) - \lambda_0^p)$
```

9. Update $p_0 \mapsto p_0 - \epsilon K_{ff}^{-1}(DR(p_0) - \lambda_0^p)$ for some small epsilon

10. Repeat steps 2 to 9 until convergence criteria is reached

5.5 Discrete Implementation details

Surfaces are discretized using triangulated meshes. The time domain is generally discretized into 10 steps. Dynamical equations for $f, p, \lambda^f, \lambda^p$ are solved using Euler's method, noting that when $\frac{d}{dt}\lambda$ is calculated to update from time $k+1$ to time k, the values of the other variables from time k are used (this is adjoint to a typical Euler step).

Images I and λ^I, velocity v_t, diffeomorphisms φ_t^{-1} and φ_{1t}^{-1}, are all sampled on a regular 3D grid x_{ijk} with fixed voxel spacing in each direction δx.

Diffeomorphisms are updated using semi-Lagrangian integration as described by Staniforth and Côté [60]. This can be written as

$$\varphi_{t_{l+1}}(x_{ijk}) = \varphi_{t_l}(x_{ijk} - \Delta t v_t(x_{ijk}))$$

with $\varphi_0 = \mathrm{Id}$, where evaluation of the image at a non-grid point is performed using trilinear interpolation. to calculate φ_{1t}^{-1}, the negative of the velocity is used, and integrated from endpoint to startpoint.

Rather than integrating the optical flow equation which may be numerically unstable, images are deformed from template directly

$$I_{t_l}(x_{ijk}) = I_0(\varphi_{t_l}^{-1}(x_{ijk}))$$

where evaluation at a non-grid point is performed using trilinear interpolation.

Derivatives are computed using centered difference

$$\nabla I(x_{ijk}) = \begin{bmatrix} \frac{1}{2\Delta x}\left(I(x_{i+1jk}) - I(x_{i-1jk})\right) \\ \frac{1}{2\Delta y}\left(I(x_{ij+1k}) - I(x_{ij-1k})\right) \\ \frac{1}{2\Delta z}\left(I(x_{ijk+1}) - I(x_{ijk-1})\right) \end{bmatrix}$$

with appropriate boundary conditions (zero normal component for images, identity for diffeomorphisms).

5.6 Image matching with geodesic parameters

Given a set of of basis functions b_i and real numbers β_i, we have the simple relationship $\delta p_0 = \sum_i b_i \delta\beta_i$, we can easily rewrite a variation in terms of p_0 as one in terms of β as shown in section 4.5.

Problem 5 (Image matching with geodesic parameters)**.** *Given a template surface parameterization f_0 and template image I_0, and a set of basis function b^i, and given some data fidelity cost $F(f_1, I_1)$ (e.g. sum of square error between I_1 and a target), and some regularization cost R (e.g. norm squared of v_0), we seek to minimize the cost*

$$C(\beta) = R(p_0) + F(f_1, I_1) \text{ or } C(p_0) = \tilde{R}(\beta) + F(f_1, I_1)$$

over β, with geodesic coordinates given by $p_0 = b^0 + \sum_{i=1}^{B} \beta^i b^i$, subject to the dynamical constraints $\frac{d}{dt} f = K_{ff} p = v(f)$, $\frac{d}{dt} I = -DIv$ and $\frac{d}{dt} p = -Dv^T(f)p$.

Statement 5 (Necessary conditions for imaging matching with geodesic parameters)**.** As in the surface matching case (statement 3), the cost function gradient with respect to β^i can be found by

$$\int_U (DR(p_0)(u) - \lambda_0^{pT}(u)) b_i(u) du$$

When using regularization cost $\tilde{R}(\beta)$ instead of $R(p_0)$, we simply have a gradient with respect to β^i of

$$D_i \tilde{R}(\beta) - \int_U \lambda_{p,0}^T(u) b_i(u) du$$

The necessary conditions are that the gradient vanishes for each i.

Proof. This proof is precisely the same as 3. □

This gives the following gradient descent algorithm

Algorithm 5.2. Image matching with geodesic parameters

1. Initialize f_0 to a template, initialize β (typically to 0)

2. Compute p_0 from β

3. Calculate f_1, p_1 by flowing $\frac{d}{dt}f = v(f)$ and $\frac{d}{dt}p = -Dv^T(f)p$ forward in time

4. Calculate $v_t(x)$, $\varphi_t^{-1}(x)$ and $\varphi_{1t}^{-1}(x)$ sampled on a dense grid

5. Calculate $DF(f_1, I_1)$

6. Initialize boundary conditions $\lambda_1^f = -D_1F^T(f_1, I_1)$, $\lambda_1^p = 0$, $\lambda_1^I = -D_2F(f_1, I_1)^T$

7. Calculate λ_t^I explicitly as $\lambda_1^I(\varphi_{1t}^{-1})|\varphi_{1t}^{-1}|$

8. Flow λ_f, λ_p backwards to $t = 0$ using (5.2) and (5.4)

9. Calculate gradient for each β_i as $\int_U DR(p_0)(u) - \lambda_0^{pT}(u))b_i(u)du$ (or $D_i\tilde{R}(\beta) - \int_U \lambda_0^{pT}(u)b_i(u)du$)

10. Update $\beta_i \mapsto \beta_i - \epsilon \left(\int_U (DR(p_0)(u) - \lambda_0^{pT}(u))b_i(u)du\right)$ (or $\beta_i \mapsto \beta_i - \epsilon \left(D_i\tilde{R}(\beta) - \int_U \lambda_0^{pT}(u)b_i(u)du\right)$) for some small ϵ

11. Repeat steps 2 to 10 until convergence criteria is reached

5.7 Examples

This work was published in [61]. Much of the text and examples below come from this paper. In every case the matching cost is sum of square error, given by

$$F(f_1, I_1) = \frac{1}{\sigma_I^2}\int_\Omega |I(x) - J(x)|^2 dx$$

for J a target image and σ_I^2 a scalar determining relative weighting between the regularization and fidelity terms.

We show the application of this methodology for embedding essentially the infinite dimensional segmentation which are now commonly available in the community to geodesic parameters of structures from three commonly available datasets. These data sets include ADNI [62] and PREDICT-HD [63], with parcellations performed with FreeSurfer and University of Iowa's quality controlled neural network technologies respectively, which are used to illustrate segmentations of subcortical structures which are now commonly available in large studies. We also include the BIOCARD [64] study of cognitive decline in preclinical normals. We focus on BIOCARD with its three temporal lobe structures analyzing hippocampus, amygdala, and entorhinal cortex segmentations, because we can carry out a complete statistical analysis demonstrating use of geodesic parameters via our parametric basis representation for statistical hypothesis testing on the etiology of the disease. We analyze the use of the geodesic parameters coupled to the prior distribution as providing automated methods for quality control, i.e. for outlier rejection and detection of common problems which occur in large neuroanatomical studies. In particular, we show example segmentations from the ADNI and PREDICT datasets that exhibit strong artifacts, demonstrating the robustness of this approach. The BIOCARD and grayscale image datasets analyzed consist of more carefully controlled data, and demonstrate typical performance.

We first apply our algorithm, with I_0 and J as a set of segmentations, to an analysis of the temporal lobe structures hippocampus, amygdala, and entorhinal cortex, in populations of patients being studied for neurodegenerative disease. The ADNI data set is also being examined in the temporal lobe and contains amygdala and hippocampus only, with segmentation coming from the automated segmentation methods. The PREDICT-HD data set allows us to focus on motor structures and includes hippocampus but not amygdala. We build a model by performing PCA on the surface mappings of left side structures from the BIOCARD dataset as described in [25]. Templates from which the mappings are performed are constructed using surface template estimation from the population of structures as in Ma [65]. The template is mapped via surface mapping onto each element in the population, from which covariances are empirically generated. We choose the first 47 eigenfunctions of the covariance for our basis functions $\{b^n\}$, accounting for 95% of the covariance's trace. We have observed that this cutoff is near the saturation point for mapping quality metrics as a function of dimension. In fact, our deformed template vertices are on average within 0.4 mm of those obtained with 160 dimensions which is well past the saturation point. This distance should be considered in the context of a typical 1mm resolution for our applications, and an average distance of 1.4 mm between vertices on the same face of our triangulated surfaces.

For each of the three datasets, a template segmentation image I_0 was constructed for each structure at the same resolution as the target segmentations. This corresponds to $0.97 \times 0.97 \times 2$ mm (with low resolution in the anterior-posterior direction) for the BIOCARD dataset, and $1 \times 1 \times 1$ mm for the other datasets. The template surfaces contouring hippocampus, amygdala, and entorhinal cortex, are depicted in the center of Figure 5.1.

Last we apply our algorithm with I_0 and J as grayscale T1 MR images. Our goal is to simultaneously identify parametric geodesic parameters of each subcortical motor structure (caudate, putamen, globus pallidus, as well as the thalamus), from the T1 image data directly. To this end we include a dataset obtained as part of a study of children with attention deficit disorder and autism spectrum disorder [30]. Patients have a mean age of 10.2 yrs, and T1-weighted 3D-volume MPRAGE coronal images were acquired from a Philips 3T Achieva MRI scanner (Best, the Netherlands) using an 8-channel head coil (TR = 7.99 ms, TE = 3.76 ms, Flip angle = 8°, voxel size = 1mm isotropic). Eight subcortical structures were segmented manually. Triangulated atlas surfaces, f_0, were constructed from the manual segmentations. We use the 30 smoothest (corresponding to the smallest eigenvalues) Laplace-Beltrami eigenfunctions per structure as a basis for dimensionality reduction, since we do not have training data for PCA. We take $b^0 = 0$ and $\sigma^2_{\alpha n}$ given by the reciprocal of one plus the magnitude squared of the eigenvalue of b^n. For typical triangulated surfaces with about 1000 vertices, this is a 100-fold reduction in the dimensionality of the problem. As compared to the full $256 \times 200 \times 256$ image this is a roughly 10000-fold reduction in the dimensionality of the problem.

5.7.1 Parameterization of target segmentations

For each subject we rigidly align the set of left temporal lobe binary segmentations to the template's space using trilinear interpolation. For the BIOCARD dataset we use manually placed landmarks to calculate this rigid transform, while for the other datasets we use unlabeled segmentation data only.

We apply our implementation of algorithm 1 using the available segmentations as image data, with $\sigma^2_I = 0.5$. The algorithm computes the transformation of state $q_1 = (f_1 = \varphi_1(f),\ I_1 = I \circ \varphi_1^{-1})$, where I is the template image, f is the template surface.

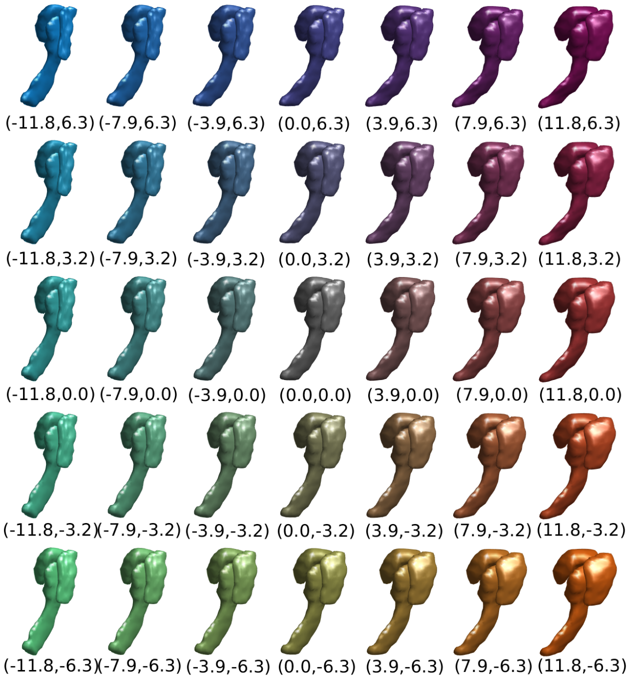


Figure 5.1: The first two dimensions of our model for hippocampus, amygdala, and entorhinal cortex are shown here. The mean shape is shown in the center in gray, and each subsequent shape is shown by taking a step of one standard deviation in the direction of the first basis function (left/right or cyan/red) or the second basis function (up/down or blue/yellow).

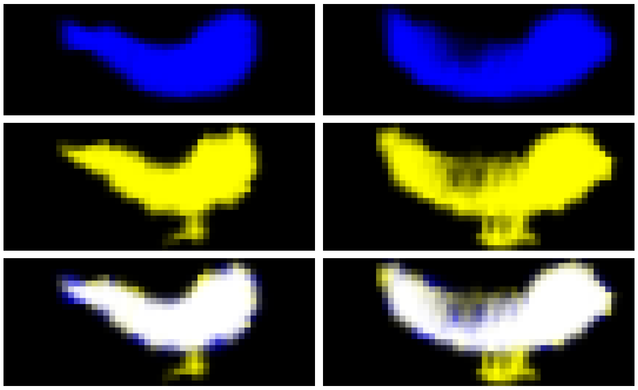


Figure 5.2: Two results (left/right) from PREDICT-HD showing a slice of the dense template image I_1 (top row), target J (second row), and both together (third row) summed in RGB space such that overlap appears white.

Hippocampus: PREDICT-HD

Figure 5.2 shows examples from the PREDICT-HD dataset of the overlap of the interior segmentation of the template hippocampus $I_0 \circ \varphi_1^{-1}$ superimposed on the target. The target segmentation is shown in yellow with the deformed hippocampus shown in blue. The region where they overlap is shown in white. The target image contains noisy components (appendages or leaks on the lateral aspect of the hippocampal body, a typical unwanted artifact in this dataset), which are filtered out by the finite dimensional representation of the shape in the geodesic coordinates of the PCA basis.

Hippocampus and amygdala: ADNI

For another pair of examples, this time from the ADNI study and including hippocampus and amygdala, Fig. 5.3 shows the segmentation data I_1 as isosurfaces. The target image contains noisy components (blobs, disconnected parts, a typical unwanted artifact in this dataset produced by FreeSurfer) which are filtered out by the finite dimensional representation the shape in the geodesic coordinates of the PCA basis. The other aspect of the state, the surfaces f_1, are shown with the first two components of their geodesic coding (bottom). Here the deformed atlas hippocampus/amygdala is shown in blue/red, and the target in yellow/cyan.

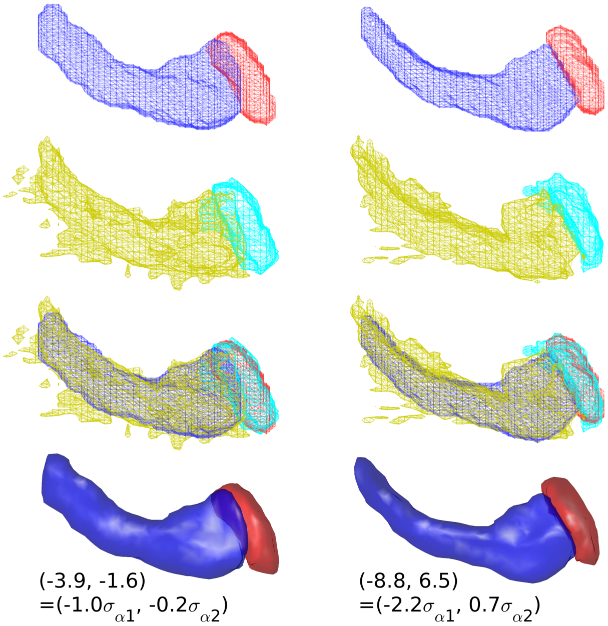


Figure 5.3: Two results (left/right) from ADNI. First row: isosurface of deformed template image I_1. Second row: isosurface of target image J. Third row: both together. Fourth row: Deformed template surfaces f_1 along with basis coefficients (in multiples of the eigenvalue standard-deviations).

Hippocampus, amygdala, and entorhinal cortex: BIOCARD

For the BIOCARD dataset we performed mappings onto the entorhinal cortex as well. One example from the BIOCARD dataset is shown in Fig. 5.4. The top left depicts the overlap of the template I_1 with the target J. In addition to the color scheme described, green depicts the deformed template entorhinal cortex, while magenta depicts the target entorhinal cortex. White depicts agreement, while other colors (RGB sums of template and target) depict misclassification errors. The middle left shows isosurfaces of I_1 compared to isosurfaces of the target J. The deformed template surfaces f_1 and first two parameters of their encoding are shown at the bottom left.

To examine the effect of the 47-dimensional prior distribution, we mapped the template onto the greater than 300 scans in all 110 subjects. To examine accuracy we computed the volume of each structure and the closeness of the mapped surfaces onto each of the three structures. Shown in the right of Fig. 5.4 are the comparisons between volumes generated by the finite-dimensional basis and the original triangulation of the quality-controlled segmentations used for studying the

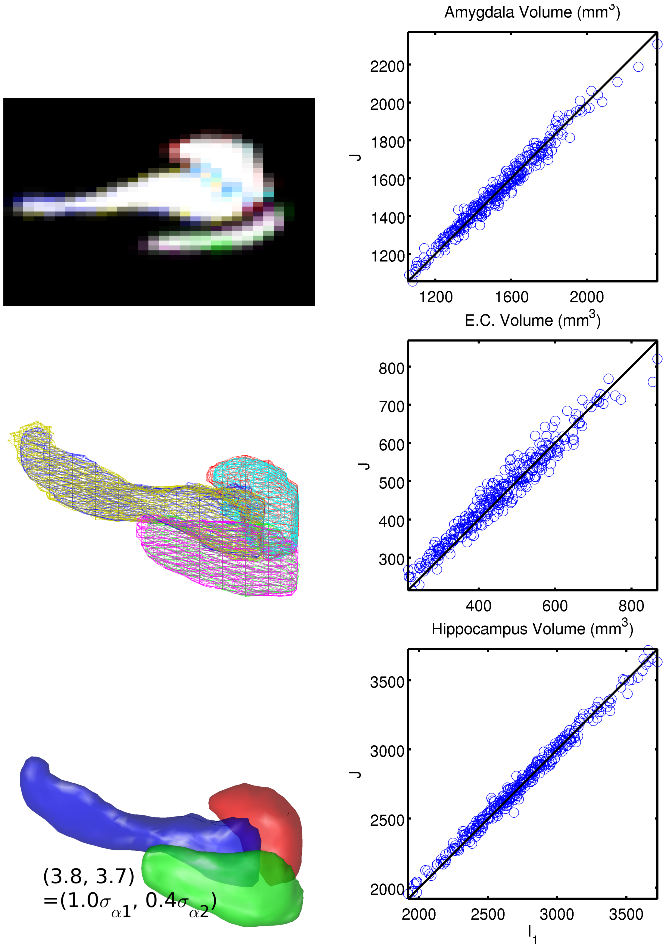


Figure 5.4: Results of mapping amygdala, entorhinal cortex, and hippocampus onto BIOCARD segmented structures. Top left: Slice through I_1 and J. Correct overlap of template and target is depicted as white, while other errors are encoded via colors as described in the text. Middle left: Isosurfaces of segmentations I_1 and J. Right: Volumes of deformed templates I_1 and targets J for the three temporal lobe structures. The identity line is shown in black.

disease already published in [66]. If these are identical then every sample will lay exactly on the black line. The greatest discrepancies occurs for the entorhinal cortex which is a very thin laminar structure, often only two voxels in laminar thickness at the original resolution, and generally not reliably characterized by volume. To view these discrepancies in context we calculated the variability which can be expected by examining volumes calculated from the dense voxel segmentation, versus calculated from the triangulated surface. We see for amygdala and hippocampus that the error is on the order of the noise of the boundary, roughly 2-4%. For entorhinal cortex we see a slight increase in discrepancy.

5.7.2 Outliers

In large studies being able to detect large discrepancies between the mapping and the target is important. Having a prior distribution allows us to test the resulting solution against the prior for outliers. The three kinds of errors which the algorithm accommodates is overlap of structures, poor initial rigid alignment, and highly distorted shapes.

Inconsistent segmentations

To study resolution of inconsistent segmentations, we use the BIOCARD dataset. Since each structure was labeled individually, there are cases where anatomical definitions overlap. We choose three examples where the amygdala and hippocampus segmentations overlap with a Dice score of 0.096, 0.072, 0.062 (roughly a 5-10 % overlap). The average overlap for the population was 0.0044, meaning that these examples overlap more than 10 times the average. Shown in the top row of Figure 5.5 are these examples. Notice that instead of a clear boundary between the yellow hippocampus, and the cyan amygdala, there exists a large zone colored green. This region corresponds to voxels which have been identified as belonging to both the hippocampus and the amygdala. The segmentations generated by the algorithm shown in the bottom row do not overlap (beyond the linear interpolation applied when generating these slices). The algorithm assigns unique labels to each voxel for three example cases, including those in the green zone. The boundary identified between the two structures

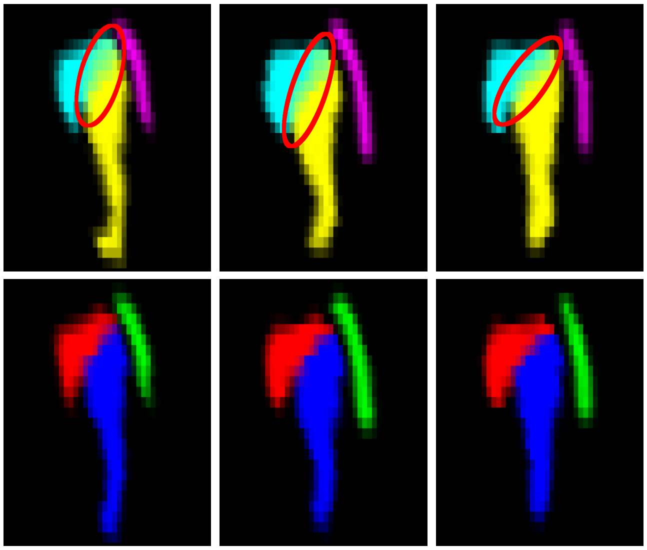


Figure 5.5: Three overlapping segmentation examples from the BIOCARD dataset. Red ellipse indicates region of overlapping labels. Top: Target. yellow/cyan/magenta: hippocampus/amygdala/entorhinal cortex. Bottom: deformed template I_1. blue/red/green: hippocampus/amygdala/entorhinal cortex.

will be the most likely, given our prior model for the shape of our temporal lobe structures.

Misoriented structures

Figure 5.6 shows examples of misregistered target segmentations depicted via isosurfaces of deformed templates I_1 and targets J which are overlayed with transparency. The misorientation becomes apparent by noticing that the body of the target hippocampus is curved the “wrong” way relative to the template. The deformed template surfaces, f_1 are shown in the bottom row, together with the first two dimensions of their coordinates, which are seen to take extreme values. In these examples, the registration is incorrectly rotated around the anterior-posterior axis by about 180 degrees. The segmentation performance is quite poor, but these cases represent quality control errors in the analysis of a large population. The prior as represented by the value of the cost $\frac{1}{2}\|p_0 - b^0\|^2_{\Sigma^{-1}_{p_0 p_0}}$ signals these outliers. Examining a set of 10 correctly registered cases and 5 misregistered cases leads to a signal to noise ratio of 2.81 between the two groups.

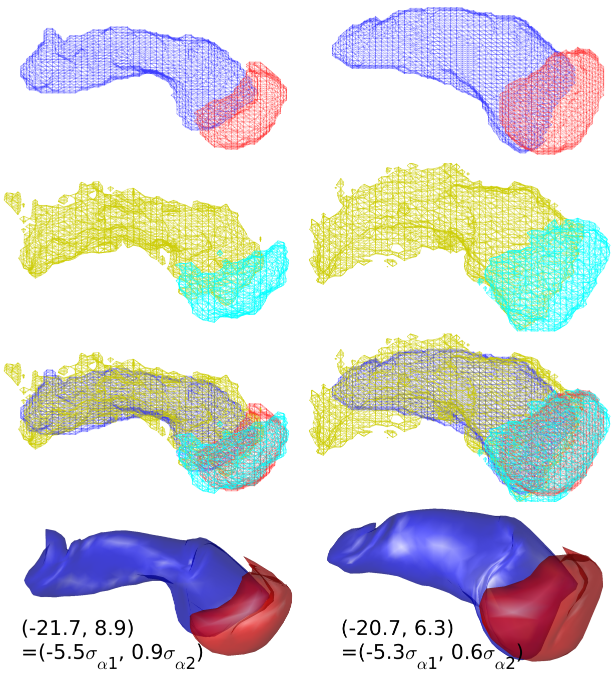


Figure 5.6: Two results (left/right) from ADNI illustrating misoriented surfaces by 180 degrees. Layout as in Fig. 5.3. Basis coefficients demonstrate the geodesic coordinates are outliers with respect to the prior.

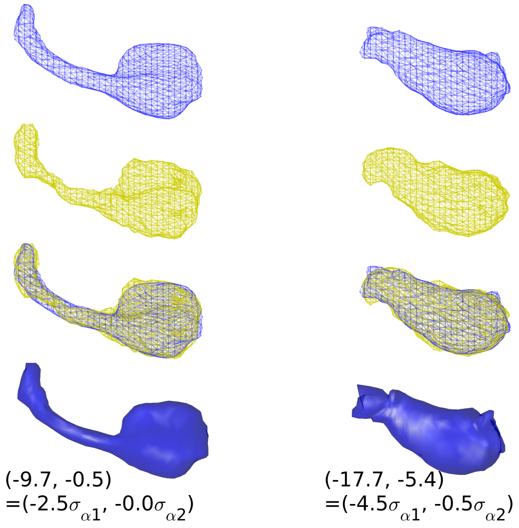


Figure 5.7: Two results (left/right) from PREDICT-HD illustrating extreme segmentations. Layout as in Fig. 5.3. Basis coefficients demonstrate the geodesic coordinates are outliers with respect to the prior.

Extreme structures

Figure 5.7 shows two extreme segmentation examples taken from the PREDICT-HD dataset. Again we use $\frac{1}{2}\|p_0 - b^0\|^2_{\Sigma^{-1}_{p_0 p_0}}$ as a signal to separate 8 normal from 5 extreme surfaces, leading to a signal to noise ratio of 1.83.

5.7.3 Reproducing statistical results

We have been using LDDMM as part of statistical analyses of large populations by quality controlling poor segmentations and triangulating them, before applying surface mapping and performing statistics. In this section we apply the parametric representation of the coordinates of the temporal lobe structures to the BIOCARD segmentations, removing the extra steps of triangulation and surface mapping. We compare the use of geodesic coordinates of the basis representation to previously published results on local (log Jacobian of φ_1 at each vertex) and volume measures [66].

The BIOCARD dataset [66] consists of the three temporal lobe structures associated to 342 brain MRI scans, consisting of 110 unique patients at up to 5 timepoints. They are divided into 3 groups: 11 clinical subjects who were diagnosed with Alzheimer's disease (AD) at the time of their last MR scan, 19 preclinical subjects who were not diagnosed with AD at the time of their last MR scan but were later diagnosed with AD, and 80 controls who did not develop AD. The temporal lobe structures were segmented and quality controlled, and a template surface generated, using procedures described in [66].

We performed statistical analyses comparing the amygdala, entorhinal cortex and hippocampus subvolumes between the two groups (controls vs. preclinical AD), modeling the local shape markers via linear and mixed effects. The model previously used in Younes et al. [67] takes the absolute volume and atrophy rate of the linear model as a function of age as different between the two groups. The mixed effects as used by Bernal-Rusiel et al. [68] corresponds to representing the noise in the measuring shape marker as corresponding to two different processes, one associated to the time series within a subject, and the second noise associated to the cross-sectional variation from subject to subject. The analysis includes age, gender and log intracranial volumes (calculated using coronal SPGR scans in FreeSurfer 5.1.0) as covariates, and computes statistics at each vertex of the triangulated template surface returning p-values corrected for multiple comparisons using permutation testing [69]. The analysis uses a mixed linear effects model for each vertex v, scan j and subject s. For this, each subject's left and right structures (controls and patients) has been registered to the template, resulting in the computation of a normalized deformation marker measuring how much expansion/atrophy at vertex v of the template surface in registering it to subject s for scan j. The raw expansion/atrophy measure is defined as the logarithm of the local expansion/reduction in surface area around the vertex, interpreted mathematically as a log-Jacobian on the template surface. This measure is then normalized for variations due to gender and intracranial volume by fitting a linear regression model that predicts the former by the latter, and taking the residual. This normalization is group independent.

We model the group variables $g(s)$ as equalling 1 if subject s belongs to the preclinical AD group and zero if the subject belongs to the control group. Our deformation marker model is given by the equation

$$y_{si}(\cdot)=(\alpha(\cdot)+\alpha'(\cdot)a_{si})+(\beta(\cdot)+\beta'(\cdot)a_{si})g_s+\epsilon_{si}(\cdot), \tag{5.5}$$

for field of markers $(\cdot)$ (either vertex or volume), MRI scan $i = 1, ...$ of subject s at age a. ϵ represents the noise, and is modeled as $\epsilon = n + \eta$ where n is a "random effect" that measures between-subject variation and η measures within-subject variation. Both processes are assumed to be centered Gaussian, with variance $\rho\sigma^2$ and σ^2, respectively. Parameters $\alpha, \alpha', \beta, \beta', \rho, \sigma$ are estimated via maximum-likelihood. We test the null hypothesis $\beta = \beta' = 0$, correcting for multiple comparisons. The test-statistic is the difference between the log-likelihoods at the maximum-likelihood estimators:

$$S(\cdot) = L^{H_1}(\cdot; \alpha, \alpha', \beta, \beta', \rho, \sigma) - L^{H_0}(\cdot; \alpha, \alpha', 0, 0, \rho, \sigma). \tag{5.6}$$

The joint test statistic is computed from Eqn. (5.5) with familywise error rate calculated by evaluating the maximum $S^* \doteq \max S(\cdot)$ over the field of statistics. The maximum value is compared to those obtained by performing the same computation a large number of times, with group labels randomly assigned to subjects. The p-value is given by the fraction of times the values of S computed after permuting the labels is larger than that obtained with the true groups. The p-values that were observed via the linear effects modeling of deformation markers are provided. The volume statistics shown provides p-values for the same linear effects model, also evaluated via permutations, in which the y is replaced by the structure volume, for which no multiple testing correction is required. More information about this type of modeling and analysis is found in the following chapter (chapter 6).

To demonstrate the validity of using the PCA basis, we performed the statistics using the original procedure already published and generated filtered surfaces based on fitting all of the subjects multiple images with the deformed templates f_1. From these surfaces, the identical procedure described above was run. Shown in Table 5.1 are a comparison of these results.

The first two columns in the table show volume and vertex results from the already published methodology. We see for the preclinical subjects only the vertex measures are signaling in the entorhinal cortex. Shown in the right two columns are p-values generated identically but from the filtered surfaces generated using the parametric basis mappings. We see almost identical results. It appears as if the volumes in ERC are on the border of discriminating. As well the vertex measures are virtually identical in p-value (.03625 versus .0291).

Table 5.1: p-values from permutation testing, showing probability of even one false positive in a multiple comparison setting.

	Original Segmentations		Parametric Segmentations	
Structure	Volume	Jacobian	Volume	Jacobian
Amygdala	0.489727	0.547545	0.395182	0.251455
Entorhinal Cortex	0.419	0.0291	0.073188	0.03625
Hippocampus	0.742636	0.132909	0.637727	0.159727

5.7.4 Grayscale Image segmentation

We choose a set of 5 skullstripped T1 images for this study. After intensity normalization and linear alignment (rigid motion and scale), each of the five atlases were mapped pairwise to the remaining four. Laplace Beltrami eigenfunctions and eigenvalues for dimensionality reduction and regularization were computed separately for each template.

Data for one example is shown in the figures below. Left side of Fig. 5.8 shows the deformed atlas as well as the target T1 images, in coronal and axial views. By design, the procedure produces an alignment which is good in the neighborhood of the subcortical structures, but poor in regions distant to them. Notice that the resolution of this dataset results in blurring of tissue boundaries on the order of 2 to 3 voxels.

The matching quality can be observed more readily by considering the alignment of gold-standard manual segmentations. The deformed atlas surfaces f_1, as well as isosurfaces of the manual segmentations for both template and target, are shown in Fig. 5.9. For more detail, the manual segmentations are shown as slices through image data in Fig. 5.8 right.

Based on the gold standard segmentations we determine accuracy using the Dice coefficient. The results for each structure, averaged across each atlas and target, are left/right caudate: 0.85/0.84, globus pallidus: 0.84/0.82, putamen: 0.87/0.88, thalamus: 0.92/0.92. The performance is greater

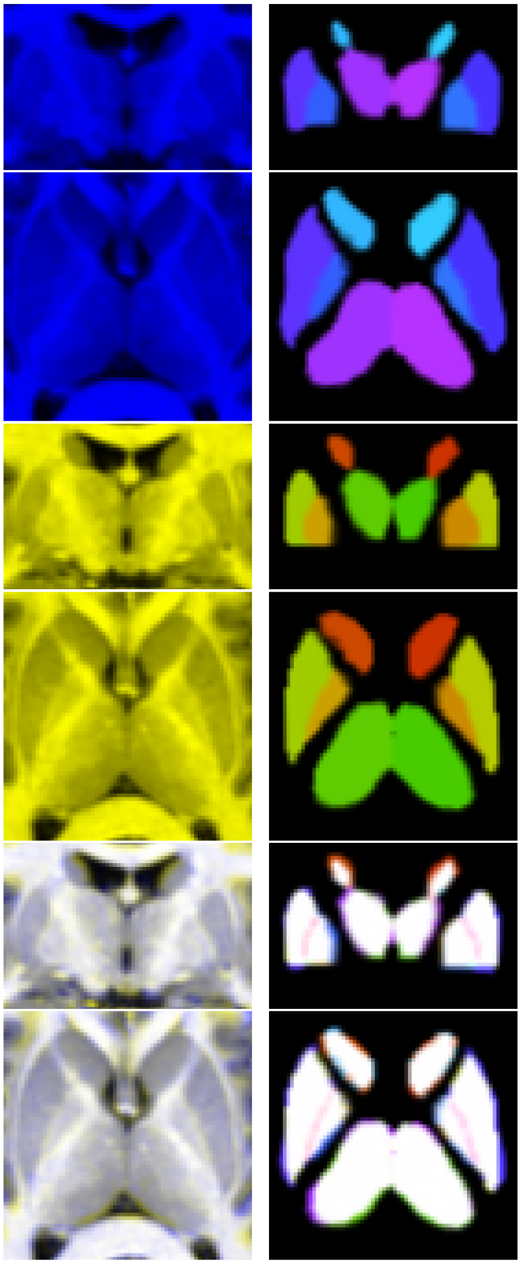


Figure 5.8: Left column: Coronal and axial slices of deformed template I_1 / target J image: blue/yellow, when summed in RGB space a correct overlap results in grayscale. The overlap is only expected to be good in neighborhood of the subcortical structures. Right column: Deformed template: cool colors, target: warm colors, when summed in RGB space a correct overlap results in white. This data was used only for evaluation, not in the mapping algorithm.

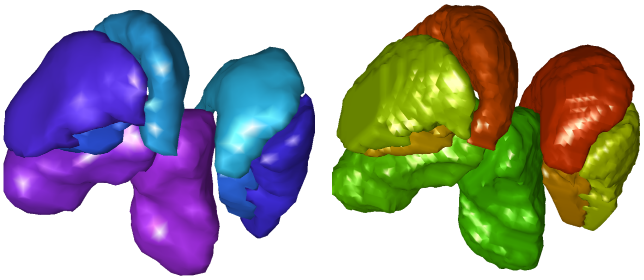


Figure 5.9: Right-inferior-anterior view of deformed template and target structures. Left: deformed atlas surfaces f_1. Right: Target segmentation isosurfaces. Data on right was used only for evaluation, not in the mapping algorithm.

than 0.8 for each structure examined, comparable to state of the art.

5.7.5 Comparison with control points methods

Control points offer an alternative to dimensionality reduction, however the surface basis function representation seems to perform better for a fixed number of dimensions.

One of our contributions is to extend the actions to states including both surface and volumes. This has allowed us to demonstrate, that for structures such as subcortical structures in the brain, efficient low-dimensional representations are obtained via Hilbert space representations based on PCA and Laplace Beltrami functions which are supported on the subcortical surface domains. This should be contrasted to other sparse methods based on point-clouds which are not as efficient for representing the local geometry of closed surface subcortical structures. To demonstrate this point, we have included an explicit result showing accuracy of mappings using only rigid alignment (6 parameters), an approach with 16 control points (48 dimensions), and our 47 dimensional prior. Control points were randomly initialized on the cortical surfaces in proportion to their surface area (4 on amygdala, 8 on hippocampus, 4 on entorhinal cortex), and forced to repel one another until they reach equilibrium, covering the entire structure as seen in Fig **??** left. We calculate the distribution of surface to surface distances between deformed template isosurfaces and target segmentation isosurfaces, for 6 examples from BIOCARD. The 95th percentile distances for amygdala

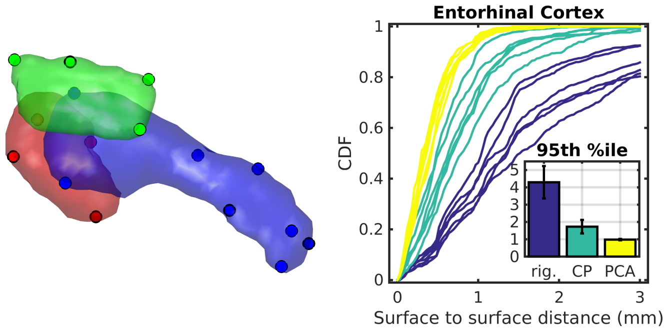


Figure 5.10: Left: The subcortical template with positioned control points (red: amygdala, blue: hippocampus, green: entorhinal cortex). Right: Distribution of entorhinal cortex surface to surface distance for rigid alignment (blue), alignment via control points (CP, cyan, 48 parameters), and alignment based on surface basis expansion (PCA, yellow, 47 parameters).

are 2.2mm (rigid) 1.5mm (control points) and 1.1mm (surface basis expansion), for hippocampus they are 2.3, 1.6, and 1.1mm respectively, and for entorhinal cortex they are 4.3, 1.7, and 0.98 mm respectively. This shows higher performance with our method in every case. Additionally, the cumulative distribution function (CDF) of surface to surface distances for entorhinal cortex is shown in Fig. 5.10 right to emphasize the consistency of our method. For example, the 95%ile distance for the worst case for the surface method is 1.1mm, but for the worst control point case it is 2.4mm, an unacceptable level of error for a structure that is only 3mm thick.

5.7.6 Discussion

Active contours or snakes as first introduced by Kass, Witkin, and Terzopoulos [43] is a a common tool in computer vision for edge detection, shape modeling, segmentation and visual tracking to name but a few applications [70, 71]. Originally models assumed explicit parameterizations, where the deformation of an initially parameterized curve is computed by minimizing an energy functional composed of a smoothing term and a term that forces the curve towards the boundary where there is a considerable change in the image intensity. Intrinsic approaches for parameterization-free deformable methods emerged using level sets to allow evolving curves to automatically merge or split [45, 51, 72], work which was later extended to deal with two-dimensional surfaces [70, 73]. Region-based and region plus edge-based based methods have been developed in 2-D [54, 74–83] and 3-D [84–88]. However, such approaches do not ensure that the evolving contour maintains important properties over time, like being simple, for example.

The general framework presented here departs significantly in that the equations of evolution are bound up with the notion that the shape is an element in a bigger metric space of shapes under flows of diffeomorphisms. This requires the vector fields satisfying the metric to have an associated number of derivatives inherited from the fact that the vector fields are in the reproducing kernel Hilbert space, and the evolution of the shape is geodesic under that metric. The focus on the indexing with respect to the submanifold boundaries implies the Hamiltonian methods of reduction which have emerged over the past 5 years in CA play a central role [89, 90].

The parametric embedding of diffeomorphisms results from the representation of the vector fields as a superposition of Green's kernels centered on the surface locations defined by the state in the Hamiltonian system, whose role is played by the surface flowing under the diffeomorphism. This gives an extremely parsimonious representation and one for which the Gaussian random field modeling is extremely efficient.

This notion of exploration of embedding of the diffeomorphic flow in a basis supported over the initial shape manifold has been explored as well in several other papers now [25, 32]. The significant departure here is to introduce an image action which combines both surface evolution and extrinsic volume evolution, allowing for a parametric decomposition of the connected subregions supported over their boundaries, while matching volume interiors. This does not assume surface to surface matching, which makes the procedure extremely robust and efficient. In this sense this circles us back to the inside-outside methods referenced above, in this topological setting that has been explored deeply by the Computational Anatomy community.

Chapter 6

Diffeomorphometry Techniques

After calculating mappings from one anatomical coordinate system to another, one typically wishes to make inferences about the biological processes at work in the population being studied. This consists of three parts, choosing an appropriate biomarker that can be computed from the mappings φ, developing a statistical model describing the distribution of these biomarkers, and performing statistical hypothesis testing on the parameters of the model a manner that controls for false positives appropriately in a multiple testing situation. For the methods developed in the following chapters, we will use one or more of these procedures to make inferences about biological processes.

6.1 Biomarkers of atrophy

Typically global biomarkers such as volume, surface area, and thickness are used to study subcortical and gyral structures. They can be easily computed from a deformed template image, when the image contains segmentations of anatomy.

To describe local changes in anatomy, biomarkers based on the Jacobian of the mappings φ are often used. When working with template surfaces, f, local analogues of volume, surface area, and thickness can be formulated.

To compute local volume change, the determinant of the Jacobian $|D\varphi|$ is used. This can be computed at each location in template space, for example at each vertex in a triangulated surface, by using (C.6) derived in section C.5.

To compute local surface area change, the determinant of the 2×2 Jacobian tangent to the template surface is used. One way this can be estimated is by simply taking the ratio of triangle areas between the deformed and the template surface. This can be interpolated onto template vertices by associating an area to each vertex, typically one third of all its neighboring triangle areas (this approach conserves total surface area).

To compute local thickness change we use the normal component of the of Jacobian. This can be interpreted as how much a thin (relative to the smoothness of the mapping) structure expands under the mapping. As such, it is appropriate for studying thin structures such as cortical gray matter. This is quantity can be computed simply as the ratio of the above.

6.2 Mixed effects modeling

Here we use mixed effects modeling to describe data involving multiple subjects, and multiple measurements of the same subject. For example several patients, each having several MRI scans.

We consider a model for each scalar valued observation (such as one Jacobian related atrophy measure described in the previous section) at a given vertex, of the following form

$$Y = X\beta + Zb + \epsilon$$

where Y is a $N \times 1$ vector of observations, X is a fixed effects design matrix, Z is a random effects design matrix, β are the fixed parameters and b are the random effects random variables, and ϵ is the noise.

We assume the noise is independent and identically distributed (IID) Gaussian with variance σ^2. We assume b is IID Gaussian with variance $\sigma^2\theta$. In the situation described above, b will have one element per subject for M subjects, and Z will be binary, having one column per subject, with ones indicating that a given observation of Y comes from this subject. X and β will describe other covariates such as age, gender or intracranial volume.

Our goal is to find maximum likelihood estimates of β, σ, θ. Additionally it will may be useful to find the expected value of b.

Statement 6 (Parameter estimation in subject/scan mixed effects models)**.** For a fixed θ, β is found by solving the linear system

$$X^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)y = X^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)X\beta$$

from which σ is found as

$$\sigma^2 = \frac{1}{N}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta) \ .$$

Parameters are estimated by maximizing the log likelihood as a function of θ

$$l(\theta) = -\frac{N}{2}\log((y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta)) + \frac{1}{2}\log|I - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T|$$

where $|\cdot|$ denotes matrix determinant and we have dropped the terms which do not depend on θ.

Proof. We write down the likelihood of our data

$$P(y|b, \beta, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}^N}\exp\left(-\frac{1}{2\sigma^2}(y - X\beta - Zb)^T(y - X\beta - Zb)\right)$$

Additionally we have

$$P(b|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2\theta}^M}\exp\left(-\frac{1}{2\sigma^2\theta}b^Tb\right)$$

We are looking for

$$\begin{aligned}
&P(y|\beta, \theta, \sigma^2) \\
&= \int P(y|\beta, b, \theta, \sigma^2)P(b|\theta, \sigma^2)db \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}^N\sqrt{2\pi\sigma^2\theta}}\exp\left(-\frac{1}{2\sigma^2}(y - X\beta - Zb)^T(y - X\beta - Zb) - \frac{1}{2\sigma^2\theta}b^Tb\right)db
\end{aligned}$$

In many cases we turn to the expectation maximization algorithm with missing data b. However, for the simple model used here this integral can be worked out exactly. In order to integrate out b we want to complete the square in the exponential C.7. We'll write out the numerator as a quadratic part in b, a linear part, and a constant part. We'll factor out $-\frac{1}{2\sigma^2}$ from each term.

$$(y - X\beta - Zb)^T(y - X\beta - Zb) + \frac{1}{\theta}b^Tb$$

$$= b^T(Z^TZ + \frac{1}{\theta})b - 2(y - X\beta)^TZb + (y - X\beta)^T(y - X\beta)$$

From the appendix C.7, we can complete the square with quadratic part $(Z^TZ + \frac{1}{\theta})$, linear part $-2Z^T(y - X\beta)$, constant part $(y - X\beta)^T(y - X\beta)$.

Plugging into the formula gives

$$\begin{aligned}(b + \frac{1}{2}(Z^TZ + \frac{1}{\theta})^{-1}(-2Z^T(y - X\beta)))^T(Z^TZ + \frac{1}{\theta})(b + \frac{1}{2}(Z^TZ + \frac{1}{\theta})^{-1}(-2Z^T(y - X\beta)))\\ + (y - X\beta)^T(y - X\beta) - \frac{1}{4}(-2Z^T(y - X\beta))^T(Z^TZ + \frac{1}{\theta})^{-1}(-2Z^T(y - X\beta))\end{aligned}$$

When we integrate over b the first term will become a normalization constant. Therefore I can write

$$P(y|\beta, \theta, \sigma^2) = \text{const} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta)\right)$$

Because this is a Gaussian distribution, we can infer that the constant term is simply related to the determinant of the covariance

$$\text{const} = \frac{|\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T|^{\frac{1}{2}}}{(2\pi\sigma^2)^{\frac{N}{2}}}$$

So that the distribution is

$$\begin{aligned}P(y|\beta, \theta, \sigma^2) = &\frac{|\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T|^{\frac{1}{2}}}{(2\pi\sigma^2)^{\frac{N}{2}}}\\ &\exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta)\right)\end{aligned}$$

The log likelihood is then

$$\begin{aligned}l(y|\beta, \theta, \sigma^2) = &-\frac{N}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log|\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T|\\ &- \frac{1}{2\sigma^2}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta)\end{aligned}$$

From the first term (that was previously ignored), we can find the posterior expected value of b, which is

$$E[b|y, \beta, \sigma, \theta] = (Z^TZ + \frac{1}{\theta})^{-1}Z^T(y - X\beta)$$

And the posterior covariance of b is

$$\text{Cov}[b|y,\beta,\sigma,\theta] = \sigma^2(Z^TZ + \frac{1}{\theta})^{-1}$$

Now for a fixed θ, it is easy to work out all the maximum likelihood estimates. We will write each as a function of θ, computing them by taking the derivative of the log likelihood and setting to zero.

First taking a gradient w.r.t. β, we have

$$X^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)y = X^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)X\beta$$

This is a linear system which can be easily solved computationally. Note that it does not depend on σ^2.

Taking the gradient w.r.t. σ^2 gives

$$\begin{aligned} &-\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta) = 0 \\ &\implies \sigma^2 = \frac{1}{N}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta) \end{aligned}$$

If we plug this expression into the log likelihood, it cancels with the last term, which becomes $-\frac{N}{2}$.

At this point, to calculate MLE's, we simply plug this in as a function of θ into a 1D optimizer (e.g. Matlab's `fminsearch`).

Note that the expression for log likelihood as a function of θ is

$$\begin{aligned} l(\theta) &= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) + \frac{1}{2}\log|I - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T| \\ &\quad - \frac{1}{2\sigma^2}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta) \\ &= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log((y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta)) + \frac{1}{2}\log|I - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T| \\ &\quad - \frac{1}{2\frac{1}{N}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta)}(y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta) \\ &= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log((y - X\beta)^T(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)(y - X\beta)) + \frac{1}{2}\log|I - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T| - \frac{N}{2} \end{aligned}$$

Note that in this distribution, the inverse covariance matrix is

$$\Sigma^{-1} = \frac{1}{\sigma^2}(\mathbf{I} - Z(Z^TZ + \frac{1}{\theta})^{-1}Z^T)$$

□

6.3 Controlling familywise error rate

When performing many hypothesis tests at once, such as one at each voxel in an image, or one at each vertex of a triangulated surface, we are interested in controlling the probability of any false positives, rather than just the probability of a false positive at each vertex. This is known as the familywise error rate, or FWER.

Suppose we have a family of N statistics U_i, $i \in \{1, \dots, N\}$, and on the first N_0 the null hypothesis is true. We will be interested in choosing a test statistic threshold, u, to control the familywise error at rate α. This means under the null hypothesis we require

$$P[\cup_{i=1}^{N_0} U_i \geq u] \leq \alpha$$

6.3.1 The Bonferroni correction

The simplest approach, and a very conservative one, is known as the Bonferroni correction. From monotonicity of probability we have

$$P[\cup_{i=1}^{N_0} U_i \geq u] \leq \sum_{i=1}^{N_0} P[U_i \geq u]$$

If we control the size of the term on the right, we will also control FWER. If we choose u such that $P[U_i \geq u] \leq \alpha/N$, then we can continue the inequalities

$$\dots \leq \sum_{i=1}^{N_0} \frac{\alpha}{N} = \frac{N_0 \alpha}{N} \leq \alpha$$

This approach simply says to divide your usual threshold for significance by N.

6.3.2 The maximum statistic

The Bonferroni approach is often far too conservative, as it does not take into account the dependence between random variables. When our random variables are related to the Jacobians of a smooth mapping, they will be far from independent, and we can generally achieve much greater power by using a different method. We observe that at least one statistic is greater than the threshold if and only if the maximum is greater than the threshold. We can express this as

$$P[\cup_{i=1}^{N_0} U_i \geq u] = P[\max_{i \in \{1,\dots,N_0\}} U_i \geq u] \leq P[\max_{i \in \{1,\dots,N\}} U_i \geq u]$$

We can therefore control the FWER by calculating the distribution of the maximum statistic (a scalar statistic, not a family of statistics). This is quite straightforward to perform analytically in the IID case, but in general we will have to turn to resampling methods such as permutation testing.

6.3.3 Permutation testing

Suppose we have a random variable X and we observe x. And suppose we have a group of transformations $\mathcal{G}$ (typically permutations), with $|\mathcal{G}|$ elements enumerated by G_i. We assume that, under the null hypothesis, the distribution of $G_i X$ is the same as the distribution of X.

Further, let's say we have a test statistic $U(X)$ and some some rejection threshold u. We want to control $P[U(X) \geq u] \leq \alpha$. Let us consider the statistic

$$T(x) = \frac{1}{|\mathcal{G}|} \sum_i I[U(G_i x) \geq U(x)]$$

for I an indicator function. This is the proportion of times that $T(G_i x)$ is bigger than $T(x)$. We want to show that the distribution $P[T(X) \leq t] \leq t$. That is, under the null hypothesis, its CDF is not above the uniform CDF.

To show this we rely on the probability integral transform. If F_X is the CDF of X, then

$$P[F_X(X^-) \geq t] \leq 1 - t$$

Here we use the convention that for some function $f(x)$, $f(x^-) = \lim_{t \to x^{-1}} f(t) = \lim_{t \to x, t < x} f(t)$. Note that in the case that X is a continuous random variable, we will have equality.

To apply this result we realize that the function $T(x)$ is related to the empirical CDF of $U(x)$ over samples from $\mathcal{G}$. We can write

$$\begin{aligned} F_{U(Gx)}(t) &= P[U(Gx) \leq t] \\ &= \frac{1}{|\mathcal{G}|} \sum_{G_i \in \mathcal{G}} I[U(G_i x) \leq t] \end{aligned}$$

so that

$$\begin{aligned} F_{U(Gx)}(t^-) &= P[U(Gx) < t] \\ &= \frac{1}{|\mathcal{G}|} \sum_{G_i \in \mathcal{G}} I[U(G_i x) < t] \end{aligned}$$

Now notice that

$$T(x) = 1 - F_{U(Gx)}(U(x)^-)$$

So using the probability integral transform we obtain

$$P[F_{U(GX)}(U(x)^-) \geq t] \leq 1 - t$$

Or in terms of $T(X)$

$$\begin{aligned} P[T(X) \leq t] &= P[1 - F_{U(GX)}(U(X)^-) \leq t] \\ &= P[1 - t \leq F_{U(GX)}(U(X)^-)] \\ &\leq 1 - t \end{aligned}$$

Therefore, if $U(X)$ is our maximum statistic, we can control the FWER at level α by choosing $1 - t \leq \alpha$. In other words, we chose our threshold t to be $\geq 1 - \alpha$. We then compute the threshold u as the $1 - t$-th quantile of the $U(Gx)$.

If we are interested in discriminating between a diseased population and a control population, the family of permutations $\mathcal{G}$ rearranges disease status between the two groups. For each permutation G_i, we calculate maximum likelihood estimates of unknown parameters as described in section 6.2, compute a test statistic at each location in our template (for example ratio of sum of square residual error under the null and alternate hypotheses), and find $U(G_i x)$ as its maximum over all vertices.

Chapter 7

Multivariate Bernoulli Matching

In this chapter we develop a technique for applying diffeomorphometry with a dataset consisting of labeled images, with a large amount of uncertainty and variability in the la belling. Do to topological differences (e.g. holes, number of connected components) among the segmentation, we will use a probabilistic atlas approach. The work here was presented at [91].

7.1 Multivariate Bernoulli fidelity term

We suppose our target image J is a vector of binary segmentations at each voxel, and our atlas image I is a probability distribution at each voxel. Letting the realizations at each target voxel x^i be independent, the likelihood can then be written as

$$L = \prod_i \prod_{c=1}^{C} I^c(x^i)^{J^c(x^i)}$$

for images with C labels (including background), with $\sum_{c=1}^{C} I^c = 1$ and $\sum_{c=1}^{C} J^c = 1$ with J^c binary.

The fidelity term is taken as the negative log likelihood

$$F(f_1, I_1) = \sum_i \sum_{c=1}^{C} -J^c(x^i)\log(I^c(x^i))$$

As is usual in these situations we define $0 \log 0 \doteq 0$.

Note that if we add an extra term that does not depend on the deformation, $J^c \log(J^c)$, this cost function will become the Kullback-Leibler divergence. This is still a meaningful cost function for the case where J is not binary, for example due to interpolations.

Problem 6 (Multivariate Bernoulli image matching)**.** *Given a template surface parameterization* f_0*, a template image* I_0 *describing the probability that each voxel takes one of* C *labels, a data fidelity cost* $F(f_1, I_1)$ *given by the negative log likelihood of a Multivariate Bernoulli distribution, and a regularization cost given by the norm squared of the geodesic coordinates, we seek to minimize the cost*

$$C(p_0) = \frac{1}{2}\|p_0\|^2_{V^*} - \sum_i \sum_{c=1}^{c} J^c(x^i)\log(I^c(x^i))$$

over the singular geodesic coordinates p_0*, subject to the dynamical constraints* $\frac{d}{dt}f = v(f), \frac{d}{dt}I = -DIv, \frac{d}{dt}p = -Dv^T(f)p$*.*

Statement 7 (Solution to the multivariate Bernoulli image matching problem)**.** We solve this problem by applying algorithm 5.1 with the appropriate cost functions R, F. To avoid avoid numerical errors we modify the cost function as follows

$$F(I_1; J, p^0) = \sum_i \sum_{c=1}^{C} -J^c(x^i)\log((1-p^0)I^c(x^i) + p^0) \tag{7.1}$$

for some small number p^0, which we take as 0.01. This is equivalent to a noise model where the label is chosen with probability p^0 from the uniform distribution, or with probability $(1-p^0)$ from the distribution in the components of I.

7.2 Probabilistic template image estimation

To begin our study of anatomical structures described through a set of binary labeled images, we will need to create a template image representing the probability of each tissue class at each voxel. The approach will be distinct from other diffeomorphic template estimation problems, such as surface template estimation described by Ma, Miller, and Younes [65], or image template estimation described by Ma et al. [92]. These start with an initial guess (or hypertemplate), and deform it to compute a template centered with respect to the population. In our setting, we will update the image intensity as well as its shape, in line with the approach used by Ashburner and Friston [93].

Given a population of target images, J^i for $i \in \{1, \dots, N\}$, we can write the likelihood of the whole population as

$$\begin{aligned}
L &= \prod_{j=1}^{N} p(J^j, \varphi^j | I) \\
&= \prod_{j=1}^{N} p(J^j | \varphi^j, I) p(\varphi^j) \\
&= \prod_{j=1}^{N} p(J^j | I(\varphi^{j,-1})) p(\varphi^j)
\end{aligned}$$

where in the last line we use the random orbit model.

Assuming an improper (uniform) prior for φ and writing p explicitly

$$\begin{aligned}
L &= \prod_{j=1}^{N} p(J^j | I(\varphi^{j,-1})) \\
&= \prod_{j=1}^{N} \prod_{c=1}^{C} \prod_{i} I^c(\varphi^{j,-1}(x^i))^{J^c(x^i)}
\end{aligned}$$

So the negative log likelihood is

$$\begin{aligned}
l &= \sum_{j=1}^{N} \sum_{c=1}^{C} \sum_{i} -J^c(x^i) \log(I^c(\varphi^{j,-1}(x^i))) \\
&\sim \sum_{j=1}^{N} \sum_{c=1}^{C} \int -J^c(x) \log(I^c(\varphi^{j,-1}(x))) dx \\
&= \sum_{j=1}^{N} \sum_{c=1}^{C} \int -J^c(\varphi^j(x)) \log(I^c(x)) |D\varphi^j(x)| dx \\
&\sim \sum_{j=1}^{N} \sum_{c=1}^{C} \sum_{i} -J^c(\varphi^j(x^i)) \log(I^c(x^i)) |D\varphi^j(x^i)| \qquad (7.2)
\end{aligned}$$

We wish to find the maximum likelihood estimate of I (or one high likelihood relative to a naive initial guess), given the constraint that $\sum_{c=1}^{C} I^c = 1$.

Problem 7 (Intensity based probabilistic template estimation)**.** *Given a population of segmentations described by binary vectors at each voxel, we wish to estimate a template I describing the probability of each tissue class at each voxel, and minimizing the negative log likelihood* (7.2) *of our population of data.*

Statement 8 (Expectation maximization iteration for intensity based probabilistic template estimation)**.** We approach the template estimation using an expectation maximization algorithm.

E: For each target j in the population perform multivariate Bernoulli matching to estimate φ^j

M: Replace the current template I as the Jacobian weighted average of each target $J^j(\varphi^j)|D\varphi^j|$

Proof. We approach this in a variational setting, enforcing the constraint that the template image must sum to one with a Lagrange multiplier α, moving to the augmented cost

$$\sum_{j=1}^N \sum_{c=1}^C \sum_i -J^c(\varphi^j(x^i)) \log(I^c(x^i))|D\varphi^j(x^i)| + \sum_i \alpha(x^i) \left(1 - \sum_{c=1}^C I^c(x^i)\right)$$

We can begin by taking a perturbation with respect to I to see

$$\frac{d}{d\epsilon} \sum_{j=1}^N \sum_{c=1}^C \sum_i -J^c(\varphi^j(x^i)) \log(I^c(x^i) + \epsilon \delta I^c(x^i))|D\varphi^j(x^i)| + \sum_i \alpha(x^i) \left(1 - \sum_{c=1}^C I^c(x^i) + \epsilon \delta I^c(x^i)\right) \Bigg|_{\epsilon=0}$$
$$= \sum_{j=1}^N \sum_{c=1}^C \sum_i -J^c(\varphi^j(x^i)) \frac{1}{I^c(x^i)} \delta I^c(x^i))|D\varphi^j(x^i)| - \sum_i \alpha(x^i) \left(\sum_{c=1}^C \delta I^c(x^i)\right)$$

For this to vanish for all perturbations we must have

$$\frac{1}{C}\alpha(x^i) = \sum_{j=1}^N -J^c(\varphi^j(x^i)) \frac{1}{I^c(x^i)} |D\varphi^j(x^i)|$$

Or more compactly

$$I^c(x^i) \sim \sum_{j=1}^N J^c(\varphi^j(x^i))|D\varphi^j(x^i)| \tag{7.3}$$

That is, I is the Jacobian weighted average of the pulled back targets, normalized to sum to 1 at each voxel.

However, this is for a fixed φ^j. In our case each φ^j is unknown and can be considered missing data. This suggests we should approach with the expectation maximization algorithm [42] over φ.

To proceed, we assume the φ are known, and write down the complete data log likelihood

$$l(J, \varphi | I) \sim \sum_{j=1}^N \sum_i \sum_{c=1}^C J^c(\varphi^j(x^i)) \log(I^c(x^i))|D\varphi^j(x^i)|$$

We wish to calculate its expected value over the conditional distribution $p(\varphi|I, J) = \frac{p(J|\varphi,I)}{\int p(J|\varphi,I)d\varphi}$. When taking the expected value we will instead use the mode approximation. i.e. for some function

f and probability density p,

$$\int f(x)p(x)dx \sim f(x^*) \text{ for } x^* = \arg\max_x p(x)$$

Therefore, we will write

$$\varphi^{j*} = \arg\max p(J^j|\varphi^j, I)p(\varphi^j|I)$$

Note that estimating this maximum can be approximated by solving the image matching problem 6. So at each iteration of the EM algorithm, we will optimize

$$\sum_{j=1}^{N}\sum_{i}\sum_{c=1}^{C} J^c(\varphi^{*j}(x^i))\log(I^c(x^i))|D\varphi^{*j}(x^i)|$$

with the constraint that our atlas probabilities sum to 1. This optimization, for a fixed $\varphi = \varphi^*$ was performed above in (7.3).

□

This gives an iterative, expectation maximization algorithm.

Algorithm 7.1. Probabilistic template image estimation

1. Rigidly align all subject segmentations
2. Initialize I to their average
3. Construct template surfaces f to support singular coordinates
4. Compute φ^{*j} for each j using image matching algorithm 5.1 (pp. 62) with cost function (7.1) (pp. 92)
5. Using φ^{*j} pull back targets to atlas coordinates, and compute Jacobian weighting (note this is the same as how λ^I is computed)
6. Update I as the Jacobian weighted average of pulled back targets.
7. Repeat steps 3 to 6 until convergence criteria is met

Iterating this EM procedure results in a sequence of template images of increasing likelihood. Often we will only use 1 iteration to create an improvement of our initial guess. This single iteration is equivalent to finding joint maximum likelihood estimates of φ^j and I and (not considering this a missing data problem).

7.3 Examples

We employ the above algorithm to studying a medial temporal lobe structures in an amnestic mild cognitive impairment (aMCI) population. The details below are from unpublished work related to [91].

7.3.1 Data

The population of aMCI patients and age matched controls enrolled in this study is described in [94]. We restate some details here. Participants completed clinical dementia rating (CDR) at baseline and underwent medial, psychiatric, neurological and neuropsychological evaluations, which included the Mini Mental Status Exam [2], Bushke Selective Reminding Test [95], the Verbal Paired Associated subtest of the Wechsler Memory Scale [96] and the Benton Visual Retention test [97]. All aMCI participants had a global CDR score of 0.5 with a sum of boxes score not exceeding 2.5 and met criteria for aMCI proposed by Petersen [98] which includes impaired memory function on testing and no decline in basic activities of daily living. All control subjects had a global CDR score of 0.

None of the aMCI participants or age-matched control subjects met criteria for dementia. Other exclusion criteria included major neurological and psychiatric disorders, head trauma with loss of consciousness, history of substance abuse or dependency, and and general contraindications to having an MRI examination (e.g. cardiac pacemaker, aneurysm coils, and claustrophobia) or taking the study medication (e.g. known sensitivity or allergies, or severe renal impairment). Participants taking anti-epileptic medications were excluded from participation in the study but use of other neuroactive medications was permitted if the participant was stable on the medication for at least 12 weeks and if the treatment regimen was not altered for the duration of the study.

The study protocol was approved by the Institutional Review Board of the Johns Hopkins Medical Institutions. All participants provided written informed consent and were paid for their participation in the study.

At the baseline evaluation 69 participants with aMCI and 24 age-matched controls met criteria for enrollment. After dropouts and image quality issues, data from 111 scans for the aMCI population and 39 scans for the control population was analyzed.

Manual Segmentations

Manual segmentations were performed, originally for the purpose of cross-participant alignment, as described in [94]. The cortical structures in the medial temporal lobe segmented were the temporal polar cortex (TPC), the entorhinal cortex (ERC), the transentorhinal region (TER, cortex lateral to ERC on both banks of the collateral sulcus), and the parahippocampal cortex. The CA1, DG/CA3 and subiculum subregions of the hippocampus were also defined in the coronal plane following landmarks described in the atlas of Duvernoy [99]. The DG/CA3 region included the CA2/CA3/CA4 and dentate gyrus subregions as these regions cannot be reliably separated on MRI scans.

Two examples of manual segmentations, one subject with aMCI and one healthy, are shown as isosurfaces in Figure 7.1.

7.3.2 Positioning Surfaces in Segmented images

The shape diffeomorphometry statistics pipeline is based on the same general protocol for subcortical structures described in [66, 100–102]. It starts from the generation of a population of surface meshes from segmentation images, from which a single template coordinate system is generated to which the shape markers are indexed by mapping of the template onto each of the target surfaces.

Since for the 300 images examined here the segmentations themselves are viewed as a noisy random observation, we now describe the procedure, distinct from “surface to surface matching” used in the above references, for positioning smooth triangulated meshes directly into the segmentations following [61].

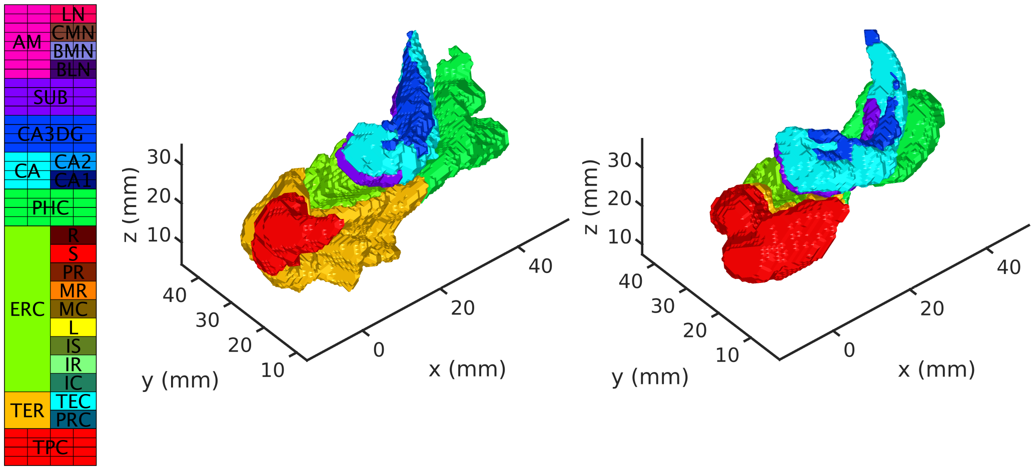


Figure 7.1: Two examples of left side manual segmentations are shown here as isosurfaces. View from superior-anterior-lateral.

We perform matching with a multivariate Bernoulli model of segmentations as described in problem 6. Two examples of surfaces positioned to match target manual segmentations are shown in Fig. 7.2. Intuitively, in regions where the atlas label is uncertain, smooth deformations are preferred over matching accuracy to eliminate overfitting of noise. In regions where the atlas label is certain, the converse is true.

We create an atlas, f_0, I_0, representative of our populations by beginning with I_0 as a simple average of rigidly aligned manual segmentations, and q_0 as a smoothed isosurface contouring the most likely label at each voxel, triangulated using a restricted Delaunay triangulation [28].

We map this atlas onto each subject in our population and update it once, defining a new I_0 as a determinant-of-Jacobian-weighted average of each manual segmentation, pulled back under the mapping. We define the updated f_0 by contouring the most likely label at each voxel as before. The atlas generation is illustrated in Fig. 7.3. This approach is simply carrying out one iteration of algorithm 7.1. Note that the atlas coordinate system is chosen in a manner which is blind to group labels.

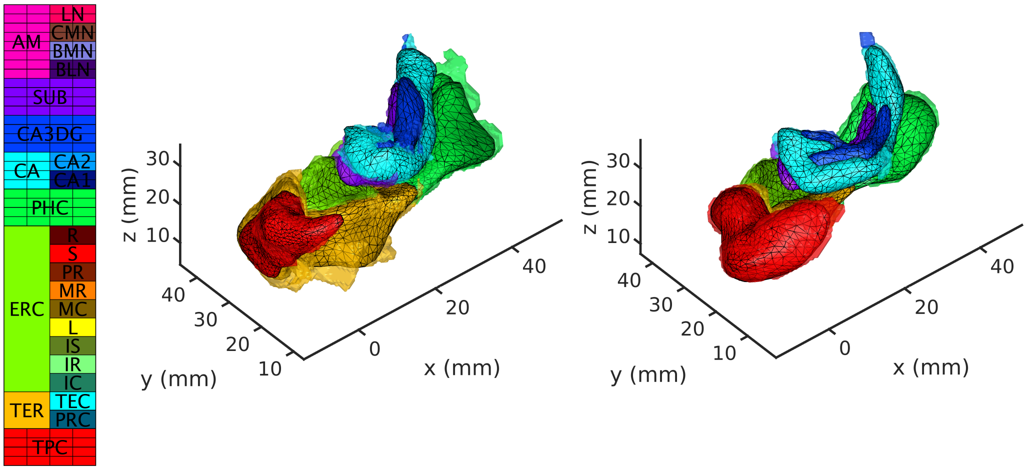


Figure 7.2: Two examples of the template (smooth triangulated surface with black edges) mapped onto the noisy segmentation data from Fig. 7.1 (semi-transparent). View from superior-anterior-lateral.

7.3.3 Shape Statistics via Diffeomorphometry

We quantify differences in shape between our two populations through the surface Jacobian, or local change in surface area. This morphometry marker can be understood as a measure of atrophy or growth. While single volume numbers can be averaged across subjects, these morphometry markers must be synchronized by building spatial correspondences across the population, requiring registration to the common template coordinates. Diffeomorphometry provides a geodesic positioning system, positioning label maps across coordinate systems. Moreover, it proves geodesic coordinates [103] which encode the shape phenotype and forms the biomarker, surface Jacobian, used in the statistical shape analysis described below.

The morphometry markers are modeled using linear mixed-effects models [66, 67, 102] at each vertex in our population atlas. Since Jacobians of diffeomorphisms are bounded below by zero, we build linear models for its logarithm. We treat age, gender, and left side versus right side, as descriptors to be corrected for, and test if there is a difference in mean between the aMCI population and the age matched controls. That is, at each vertex we consider the log surface Jacobian to be a

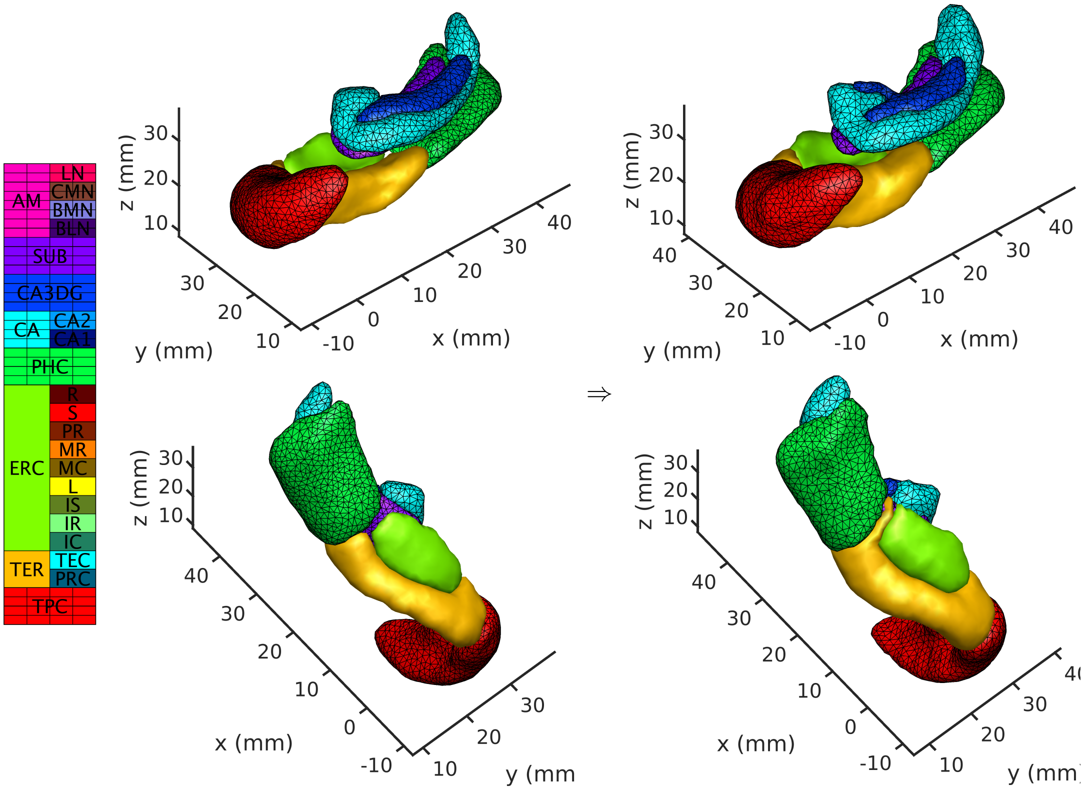


Figure 7.3: Surfaces contouring the structures of our probabilistic atlas. Left: initial atlas, Right: improved resolution atlas. Top: view from superior-anterior-lateral, Bottom: view from inferior-posterior-medial.

random variable of the form

$$\log\left(\text{Surface Jacobian}\right) = a + a'(\text{disease}) + b(\text{left}) + c(\text{female}) + d(\text{age}) + \text{patient noise} + \text{scan noise} \tag{7.4}$$

where the patient noise term is our random effect and the other terms are fixed effects. We test if there is evidence to reject the null hypothesis that $a' = 0$.

The parameters a, a', b, c, d were estimated at each vertex by maximum likelihood as shown in section 6.2, and a test statistic was formed at each vertex by computing the ratio of the sum of square fitting error between the null and alternate model. If the noise is Gaussian, this corresponds to a most powerful likelihood ratio test statistic, but the Gaussian assumption is not required in our nonparametric tests.

Permutation testing was performed with 50 000 permutations to determine statistical significance. Significant regions with familywise error rate (FWER) controlled at 5% were calculated with permutation testing as described in [69].

7.3.4 Results and Discussion

Regions where we have evidence to reject the null hypothesis with FWER controlled at 5% are contoured in Fig. 7.4. One notes a large region of atrophy detected in the trans entorhinal region, at the level of the rostral entorhinal cortex, surrounding the collateral sulcus.

The extent of atrophy is also quantified in the FWER controlled regions in Fig. 7.4. The exponential of the parameter a' from equation (7.4) is plotted at every vertex. Up to 6% atrophy of the aMCI group relative to the age matched controls is observed around the collateral sulcus.

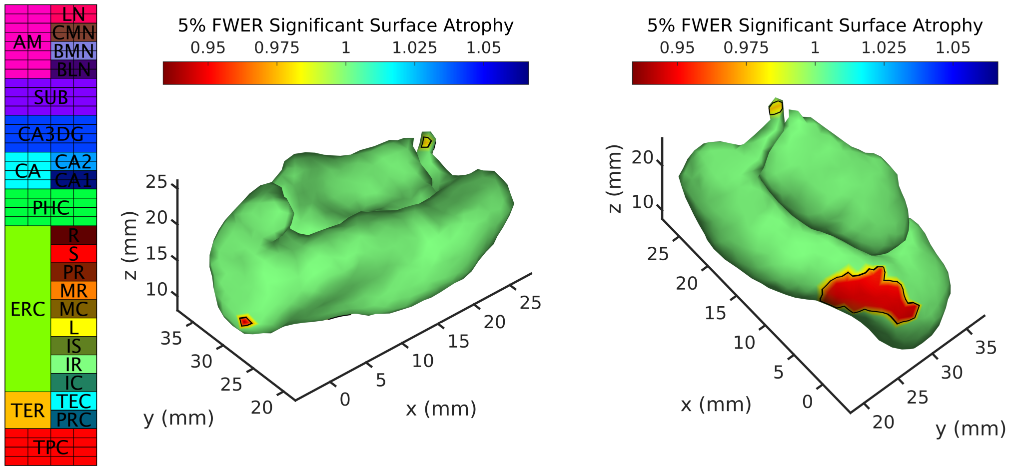


Figure 7.4: Surface atrophy in regions of change for FWER at 5%.

Chapter 8

Diffeomorphic template estimation with singular geodesic coordinates

In addition to the intensity based, probabilistic template estimation of section 7.2, we also present a template estimation method which determines an optimal deformation of an initial guess, using singular geodesic coordinates to construct the deformation. The result is something of a hybrid between [92] and [65] because it uses deforming surfaces, as well as Jacobian weighted averages of images.

8.1 Cost function and dynamical constraints

We estimate a template centered to a given population as a deformation of an initial guess, or hypertemplate, consisting of an image I^0 and a surface parameterization f^0. There are two geodesic trajectories linking the initial guess to each of the N members of the population, images J^i. One from hypertemplate to the template we wish to estimate, and one from this template to each subject. We will use a superscript 0 for variables describing the trajectory from hypertemplate to template, and and a superscript $i \in \{1, \ldots, N\}$ for variables describing the trajectory between the template and the ith target. Because these trajectories are continuous, we will need $f^0_1 = f^i_0$ and $I^0_1 = I^i_0$ for all i. But there is no requirement of continuity for the coordinates p, meaning p_1^0 is in generally not equal to p_0^i.

Problem 8 (Diffeomorphic template estimation with singular geodesic coordinates)**.** *Given a hypertemplate surface parameterization, f_0^0, and hypertemplate image I_0^0, and given a set of regularization functions $R^i(p_i^0, f_i^0)$ one for the hypertemplate and one for each of the N members of the population being studied (e.g. norm squared of each geodesic coordinate), and given a set of data fidelity functions $F^i(I_1^i)$, one for each member of the population (e.g. sum of square error), we seek to minimize the cost*

$$C(p_0^0, p_0^i) = R^0(p_0^0, f_0^0) + \sum_{i=1}^{N} R^i(p_0^i, f_0^i) + F^i(f_1^i, I_1^i)$$

over all the geodesic coordinates p_0^i for $i \in \{0, \dots, N\}$, subject to the dynamical constraints $\frac{d}{dt}f^i = v^i(f^i)$, $\frac{d}{dt}I^i = -DI^i v^i$, $\frac{d}{dt}p^i = -Dv^{iT}(f^i)p^i$.

Notice that in this case our regularization functions are shown to depend explicitly on the surfaces f^i

As before we enforce dynamics with Lagrange multipliers, and write out the augmented cost:

$$\begin{aligned}
&R^0(p_0^0, f_0^0) + \sum_{i=1}^{N} R^i(p_0^i, f_0^i) + F^i(f_1^i, I_1^i) \\
&\quad + \int_0^1 (\lambda^{0,f} | \frac{d}{dt} f^0(u) - \int_U K(f^0(u), f^0(u'))p^0(u')du') \\
&\quad + (\lambda^{0,I} | \frac{d}{dt} I^0(x) + DI^0(x) \int_U K(x, f^0(u'))p^0(u')du') \\
&\quad + (\lambda^{0,p} | \frac{d}{dt} p^0(u) + \int_U \nabla_1 K(f^0(u), f^0(u'))p^{0T}(u)p^0(u')du') \\
&\quad + \sum_{i=1}^{N} (\lambda^{i,f} | \frac{d}{dt} f^i(u) - \int_U K(f^i(u), f^i(u'))p^i(u')du') \\
&\quad + (\lambda^{i,I} | \frac{d}{dt} I^i(x) + DI^i(x) \int_U K(x, f^i(u'))p^i(u')du') \\
&\quad + (\lambda^{i,p} | \frac{d}{dt} p^i(u) + \int_U \nabla_1 K(f^i(u), f^i(u'))p^{iT}(u)p^i(u')du')dt
\end{aligned}$$

8.2 Necessary conditions

The dynamics of the Lagrange multipliers will be exactly the same same those for the simpler image matching problem, described in (5.2), (5.3), (5.4). So we'll only consider boundary terms, noting that the λ may not be continuous.

Statement 9 (Necessary conditions for diffeomorphic template estimation with singular geodesic coordinates)**.** Dynamical equations describing stationary solutions for each λ on the interior of each interval $t \in (0, 1)$ are given in statement 4.

Boundary conditions are given by

$$\lambda_1^{i,f} = -D_1 F^{iT}(f_1^i, I_1^i), \qquad \lambda_1^{0,f} = -D_2 R^{iT}(p_0^i, f_0^i) + \sum_{i=1}^{N} \lambda_0^{i,r}$$

$$\lambda_1^{i,I} = -D_2 F^{It}(F_1^I, I_1^i), \qquad \lambda_1^{0,I} = \sum_{i=1}^{N} \lambda_0^{i,I}$$

$$\lambda_1^{i,p} = 0, \qquad \lambda_1^{0,p} = 0$$

$$\lambda_0^{i,p} = D_1 R^{iT}(p_0^i, f_0^i), \qquad \lambda_0^{0,p} = D_1 R^{0T}(p_0^0, f_0^0)$$

Proof. The conditions are derived below using a variational approach, arising from boundary terms when applying integration by parts.

8.2.1 Perturbation with respect to f

We consider $f^0 \mapsto f^0 + \epsilon \delta f^0$ with $\delta f^0 = 0$. And we consider $f^i \mapsto f^i + \epsilon \delta f^i$ for $i \in \{1, \ldots, N\}$. We will require that $\delta f_0^i = \delta f_{1^-}^0$ (each trajectory is continuous). Note that $\delta f_0^0 = 0$.

For f^0

$$(\lambda_{1^-}^{0,f} | \delta f_{1^-}^0) - (\lambda_0^{0,f} | \delta f_0^0)$$

The second term is zero because we have a fixed hypertemplate.

For f^i

$$(D_2 R^{iT}(p_0^i, f_0^i) | \delta f_0^i) + (D_1 F^{iT}(f_1^i, I_1^i) | \delta f_1^i) + (\lambda_1^{i,f} | \delta f_1^0) - (\lambda_0^{i,f} | \delta f_0^i)$$

Putting this together gives

$$\lambda_1^{i,f} + D_1 F^{iT}(f_1^i, I_1^i) = 0$$

$$\lambda_{1^-}^{0,f} + \sum_{i=1}^{N} -\lambda_0^{i,f} + D_2 R^{iT}(p_0^i, f_0^i) = 0$$

8.2.2 Perturbation with respect to I

We consider $I^0 \mapsto I^0 + \epsilon \delta I^0$ with $\delta I^0 = 0$. And we consider $I^i \mapsto I^i + \epsilon \delta I^i$ for $i \in \{1, \ldots, N\}$. We will require that $\delta I_0^i = \delta I_{1^-}^0$ (each trajectory is continuous). Note that $\delta I_0^0 = 0$.

For I^0

$$(\lambda_{1^-}^{0,I} | \delta I_{1^-}^0) - (\lambda_0^{0,I} | \delta I_0^0)$$

the second term is zero because we consider a fixed hypertemplate.

And for I^i

$$(D_2 F^{iT}(f_1^i, I_1^i) | \delta I_1^i) + (\lambda_{1^-}^{i,I} | \delta I_{1^-}^{i,I}) - (\lambda_0^{i,I} | \delta I_0^{i,I})$$

Putting this together gives

$$\lambda_1^{i,I} = -D_2 F^{iT}(f_1^i, I_1^i)$$

and

$$\lambda_{1^-}^{0,I} - \sum_{i=1}^{N} \lambda_0^{i,I} = 0$$

8.2.3 Perturbation with respect to p

We consider $p^0 \mapsto p^0 + \epsilon \delta p^0$ with $\delta p^0 = 0$. And we consider $p^i \mapsto p^i + \epsilon \delta p^i$ for $i \in \{1, \ldots, N\}$.

We consider $f^0 \mapsto f^0 + \epsilon \delta f^0$ with $\delta f^0 = 0$. And we consider $f^i \mapsto f^i + \epsilon \delta f^i$ for $i \in \{1, \ldots, N\}$. We will require that $\delta f_0^i = \delta f_{1^-}^0$ (each trajectory is continuous). We have

$$(D_1 R^{0T}(p_0^0, f_0^0) | \delta p_0^0) + (\lambda_{1^-}^{0,p} | \delta p_{1^-}^0) - (\lambda_0^{0,p} | \delta p_0^0)$$

and

$$\sum_{i=1}^{N} (D R^{iT}(p_0^i, f_0^i) | \delta p_0^i) + (\lambda_1^{i,p} | \delta p_1^i) - (\lambda_0^{i,p} | \delta p_0^i)$$

Recall that unlike for I and f, we do not require $\delta p_1^0 = \delta p_0^i$.

Putting this together gives

$$D_1 R^{0T}(p_0^0, f_0^0) - \lambda_0^{0,p} = 0$$

As before, with all the other constraints obeyed, this term will be the gradient with respect to p_0^0.

We also have the conditions

$$\lambda_{1^-}^{0,p} = 0$$

and

$$D_1 R^{iT}(p_0^i, f_0^i) - \lambda_0^{i,p} = 0$$

where as before this term is interpreted as the gradient of the cost with respect to each p^i.

And finally we have the usual endpoint boundary term

$$\lambda_1^{i,p} = 0$$

8.2.4 Gradient of the cost function

We can use the interpretation

$$\frac{d}{dp^i} C = D_1 R^{iT}(p_0^i, f_0^i) - \lambda_0^{i,p}$$

so that the gradient vector is

$$K_{f_0^i f_0^i}^{-1}(D_1 R^{iT}(p_0^i, f_0^i) - \lambda_0^{i,p})$$

And at the hypertemplate

$$\frac{d}{dp^0} C = d_1 R^{0T}(p_0^i, f_0^i) - \lambda_0^{0,p}$$

so that the gradient vector is

$$K_{f_0^0 f_0^0}^{-1}(D_1 R^{0T}(p_0^0, f_0^0) - \lambda_0^{0,p})$$

8.3 Algorithm for diffeomorphic template estimation with singular geodesic coordinates

Algorithm 8.1. Diffeomorphic template estimation with singular geodesic coordinates

1. Initialize f_0^0 and I_0^0 to a hypertemplate, initialize p_0^0 and p_0^i for $i \in \{1, \ldots, N\}$ (typically to 0)
2. Calculate f_t^0, p_t^0 by flowing $\frac{d}{dt} f^0 = v^0(f^0)$, $\frac{d}{dt} p^0 = -Dv^{0T}(f^0)$ forward in time for $t \in [0, 1)$
3. Calculate f_t^i, p_t^i for $i \in \{1, \ldots, N\}$ by flowing $\frac{d}{dt} f^i = v^i(f^i)$, $\frac{d}{dt} p^i = -Dv^{iT}(f^i)$ forward in time for $t \in [0, 1]$
4. Calculate $v_t^0(x)$, $\varphi_t^{0,-1}(x)$ and $\varphi_{1t}^{0,-1}(x)$ and $I_t^0(x)$ for $t \in [0, 1), x \in \Omega$
5. Calculate $v_t^i(x)$, $\varphi_t^{i,-1}(x)$ and $\varphi_{1t}^{i,-1}(x)$ and $I_t^i(x)$ for $i \in \{1, \ldots, N\}$ for $t \in [0, 1], x \in \Omega$
6. Calculate $D_1F^i(f_1^i, I_1^i)$,$D_2F^i(f_1^i, I_1^i)$, $D_1R^i(p_0^i, f_0^i)$, $D_2R^i(p_0^i, f_0^i)$ and $D_1R^0(p_0^0, f_0^0)$
7. Initialize $\lambda_1^{i,f} = -D_1F^{iT}(f_1^i, I_1^i)$, $\lambda_1^{i,I} = -D_2F^{iT}(f_1^i, I^i)$, $\lambda_1^{i,p} = 0$
8. Calculate $\lambda_t^{i,I} = \lambda_1^{i,I}(\varphi_{1t}^{i,-1})|D\varphi_{1t}^{i,-1}|$ explicitly for $t \in [0, 1]$.
9. Flow $\lambda_t^{1,f}$, $\lambda_t^{1,p}$ backwards from $t = 1$ to $t = 0$ using (5.2), (5.4).
10. Calculate gradient vector $K^{-1}_{f_0^i f_0^i}(D_1R^{iT}(p_0^i, f_0^i) - \lambda_0^{1,p})$
11. Initialize $\lambda_1^{0,f} = \sum_{i=1}^{N} \lambda_0^{i,I} - D_2R^{iT}(p_0^i, f_0^i)$, $\lambda_1^{0,I} = \sum_{i=1}^{N} \lambda_0^{i,I}$ and $\lambda_1^{0,p} = 0$
12. Calculate $\lambda_t^{0,I} = \lambda_1^{0,I}(\varphi_{1t}^{0,-1})|D\varphi_{1t}^{0,-1}|$ explicitly for $t \in [0, 1]$
13. Flow $\lambda_t^{0,f}$, $\lambda_t^{0,p}$ backwards from $t = 1$ to $t = 0$ using (5.2), (5.4)
14. Calculate gradient vector $K^{-1}_{f_0^0 f_0^0}(D_1R^{0T}(p_0^0, f_0^0) - \lambda_0^{0,p})$
15. Update $p_0^0 \mapsto p_0^0 + \epsilon K_{f_0^0 f_0^0}(D_1R^{0T}(p_0^0, f_0^0) - \lambda_0^{0,p})$ for some small ϵ
16. Update $p_0^i \mapsto p_0^i + \epsilon K_{f_0^i f_0^i}(D_1R^{iT}(p_0^i, f_0^i) - \lambda_0^{i,p})$ for $i \in \{1, \ldots, N\}$ for some small ϵ
17. Repeat steps 2 to 16 until convergence criteria is met.

8.4 Examples

Here we show examples applying the template estimation algorithm to the entorhinal cortex and trans entorhinal cortex for the purpose of quantifying atrophy in early Alzheimer's disease.

8.4.1 Entorhinal cortex and trans entorhinal cortex

We consider a dataset consisting of manually segmented left entorhinal and trans-entorhinal cortex, in subjects without a disconnection in their collateral sulcus, from baseline scans in the ADNI dataset. A total of 34 subjects were found matching these criteria.

To estimate a hypertemplate, each segmentation was rigidly aligned and averaged. An isosurface, f_0^0, was generated by thresholding at the value 0.5. The hypertemplate image, I_0^0 was generated by filling each voxel with its fraction inside the surface.

Algorithm 8.1 was run for 2000 iterations of gradient descent, with a Gaussian kernel with standard deviation 6.0 mm. The results are shown in Fig. 8.1. To position the hypertemplate and template relative to the population, the coefficient with respect to the first two principal components of initial momentum from the template for the population were used to embed in a 2D plane. For the hypertemplate, the negative of its final momentum vector ($-p_1^0$) was used.

In the deformation from hypertemplate to template, note that changes in thickness, curvature, and boundary extent occur.

8.4.2 Entorhinal and trans entorhinal cortex: 2 channels

In a second study of the entorhinal and trans entorhinal cortex, manual segmentations of baseline scans from a population of 80 subjects from the BIOCARD population were used to estimate a template.

In this case we use $C = 2$, a 2 channel image, with one channel as the ERC segmentation and one as the TEC segmentation. This approach may allow a more accurate description of the boundary between them.

Fig. 8.2 shows results as described above for left entorhinal and transentorhinal cortex, while

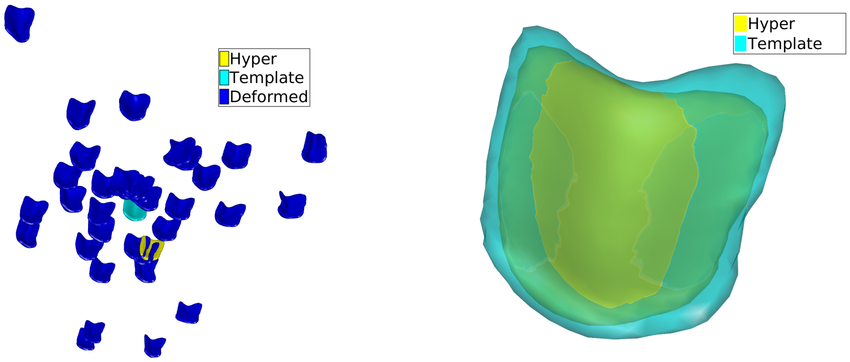


Figure 8.1: Hypertemplate (yellow) and its deformation to template (cyan) are shown. Left shows population structures (blue) laid out in a 2D plane according to the first two PCA components of p_0^i. Right shows hypertemplate and template overlayed so their differences can be inspected. The surface is oriented as though looking inside the head of a subject facing away. Left: lateral, right: medial, up: rostral, down: caudal. The superior surface is visible.

Fig. 8.3 shows the right side.

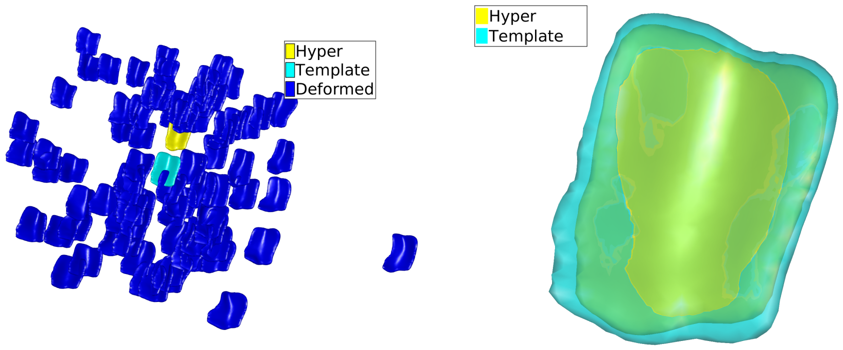


Figure 8.2: Left hypertemplate (yellow) and its deformation to template (cyan) are shown. Left shows population structures (blue) laid out in a 2D plane according to the first two PCA components of p_0^i. Right shows hypertemplate and template overlayed so their differences can be inspected. The surface is oriented as though looking inside the head of a subject facing away. Left: lateral, right: medial, up: rostral, down: caudal. The superior surface is visible.

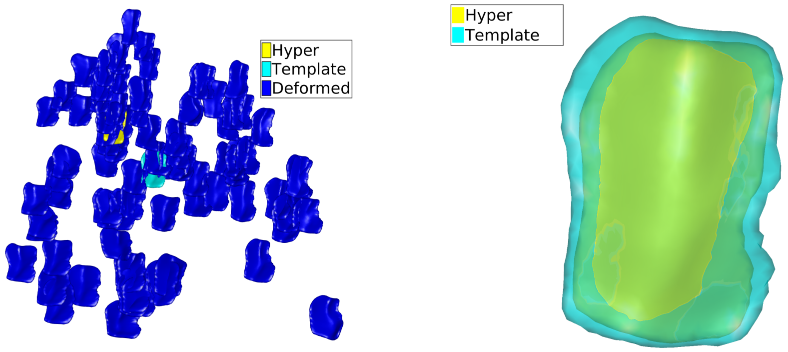


Figure 8.3: Right hypertemplate (yellow) and its deformation to template (cyan) are shown. Left shows population structures (blue) laid out in a 2D plane according to the first two PCA components of p_0^i. Right shows hypertemplate and template overlayed so their differences can be inspected. The surface is oriented as though looking inside the head of a subject facing away. Left: medial, right: lateral, up: rostral, down: caudal. The superior surface is visible.

Chapter 9

Longitudinal image matching with singular geodesic coordinates

It is natural to use the geodesic trajectories we have been considered to model the passage of actual time. Here we develop an algorithm to do just that, mapping a template through a timeseries of images using two geodesic trajectories. One from template to baseline, and one from baseline through each follow up. These trajectories can be thought of as analogous to "intercept" and "slope" in simple linear regression.

Longitudinal FreeSurfer [33, 104] addresses this issue with a common initialization of optimization problems for each scan in a timeseries. It avoids modeling any growth or atrophy process with the intention of avoiding bias by privileging a given (e.g. baseline) scan, and to allow the capture of sudden changes.

Several models for growth and atrophy scenarios using flows of diffeomorphisms are discussed in [105], with a focus on modeling populations of timeseries, and describing relationships of a given growth process to a typical one. A general treatment of regression on Reimannian manifolds is described in [106], with more complex statistical processes are described through hierarchical geodesic models in [107]. In [108], several parameterizations of these flows are considered, including piecewise geodesic (also used in [105]), spline based, and geodesic shooting (in order of increasing regularity in time).

We will use this technique for the purpose of filtering out inconsistencies over time from manual segmentations that are challenging to perform accurately. As such, over regularization is not a concern, and the proposed parameterization (geodesic shooting) is appropriate.

9.1 Cost function and dynamical constraints

Since two geodesic trajectories are used to map our template to each member of the population, we will use a similar notation to the previous chapter. Superscript 0 will be used to represent quantities along the trajectory from template to baseline, and superscript 1 will be used to represent quantities along the trajectory from template through each follow up.

The first trajectory will be indexed in time using $t \in [0, 1]$, while the second will be indexed using real time, with measurements at times t_i for $i \in \{1, \dots, N\}$ which we will denote J_{t_i}.

As before we will have continuity in f and I, meaning $f^0_1 = f^1_{t_1}$ and $I^0_1 = I^1_{t_1}$, and there will be no such necessity on the variables p.

Problem 9 (Longitudinal image matching with singular geodesic coordinates)**.** *Given a template surface parameterization f^0_0 and image I^0_0, and given a pair of regularization functions R^0 and R^1 (e.g. norm squared of each geodesic coordinate), and a set of data fidelity functions F^i, we wish to minimize the cost*

$$C(p^0_0, p^1_0) = R^0(p^0_0, f^0_0) + R^1(p^1_{t_1}, f^1_{t_1}) + \sum_{i=1}^{N} F^i(f_{t_i}, I_{t_i})$$

over the variables p^0 and p^1, subject to the dynamical constraints $\frac{d}{dt} f^i = v^i(f^i)$, $\frac{d}{dt} I^i = -DI^i v^i$, $\frac{d}{dt} p^i = -Dv^{iT}(f^i) p^i$.

As usual we will enforce the dynamical constraints using Lagrange multipliers, and show that the augmented cost function is stationary to perturbations

$$\begin{aligned} &R^0(p^0_0, f^0_0) + R^1(p^1_{t_1}, f^1_{t_1}) + \sum_{i=1}^{N} F(f^1_{t_i}, I^1_{t_i}) \\ &\quad + \int_0^1 (\lambda^{0,f} | \frac{d}{dt} f^0(u) - \int_U K(f^0(u), f^0(u')) p^0(u') du') \end{aligned}$$

$$
\begin{aligned}
&+ (\lambda^{0,I} | \frac{d}{dt} I^0(x) + DI^0 \int_U K(x, f^0(u')) p^0(u') du') \\
&+ (\lambda^{0,p} | \frac{d}{dt} p^0(u) + \int_U \nabla_1 K(f^0(u), f^0(u')) p^{0T}(u) p^0(u') du') dt \\
&+ \int_{t_1}^{t_N} (\lambda^{1,f} | \frac{d}{dt} f^1(u) - \int_U K(f^1(u), f^0(u')) p^1(u') du') \\
&+ (\lambda^{1,I} | \frac{d}{dt} I^1(x) + DI^0 \int_U K(x, f^1(u')) p^1(u') du') \\
&+ (\lambda^{1,p} | \frac{d}{dt} p^1(u) + \int_U \nabla_1 K(f^1(u), f^1(u')) p^{1T}(u) p^1(u') du') dt
\end{aligned}
$$

9.2 Necessary conditions

The dynamics of the Lagrange multipliers will be exactly the same same those for the simpler image matching problem, described in (5.2), (5.3), (5.4). So we'll only consider boundary terms, noting that the λ may not be continuous.

Statement 10 (Necessary conditions for longitudinal image matching with singular geodesic coordinates)**.** Dynamical equations describing stationary solutions for each λ on the interior of each interval $t \in (0, 1)$ are given in statement 4.

Boundary conditions are given by

$$
\begin{aligned}
\lambda^{1,f}_{t_N} &= -D_1 F^{iT}(f^1_{t_N}, I^1_{t_N}) \\
\lambda^{1,f}_{t_i^-} &= \lambda^{1,f}_{t_i} - D_1 F^{iT}(f^1_{t_i}, I^1_{t_i}) \\
\lambda^{0,f}_1 &= \lambda^{1,f}_{t_1^{-1}} - D_2 R^{1T}(p^1_{t_1}, f^1_{t_1}) - D_1 F^{1T}(f^1_{t_1}, I^1_{t_1})
\end{aligned}
$$

and

$$
\begin{aligned}
\lambda^{1,f}_{t_N} &= -D_1 F^{iT}(f^1_{t_N}, I^1_{t_N}) \\
\lambda^{1,f}_{t_i^-} &= \lambda^{1,f}_{t_i} - D_1 F^{iT}(f^1_{t_i}, I^1_{t_i})
\end{aligned}
$$

and

$$
\begin{aligned}
\lambda^{1,p}_{t_N} &= 0 \\
\lambda^{1,p}_{t_i^-} &= \lambda^{1,p}_{t_i}
\end{aligned}
$$

$$\lambda^{1,p}_{t_1} = D_1 R^{1T}(p^1_{t_1}, f^1_{t_1})$$

$$\lambda^{0,p}_{1^-} = 0$$

$$\lambda^{0,p}_{0} = D_1 R^{0T}(p^0_{0}, f^0_{0})$$

Proof. The conditions are derived below using a variational approach, arising from boundary terms when applying integration by parts.

9.2.1 Perturbation with respect to f

We consider $f^0 \mapsto f^0 + \epsilon\delta f^0$ with $\delta f^0 = 0$. And we consider $f^1 \mapsto f^i + \epsilon\delta f^1$. We will require that $\delta f^1_0 = \delta f^0_{1^-}$ (each trajectory is continuous). Note that $\delta f^0_0 = 0$.

For f^0 we have the terms

$$(\lambda^{0,f}_{1^-}|\delta f^0_{1^-}) - (\lambda^{0,f}_{0}|\delta f^0_{0})$$

The second term is zero because we are considering a fixed template.

For f^1

$$\begin{aligned}&(D_2 R^{1T}(p^1_{t_1}, f^1_{t_1})|\delta f^1_{t_1}) + (D_1 F^{1T}(f^1_{t_1}, I^1_{t_1})|\delta f^1_{t_1})\\ &\quad + \sum_{i=1}^{N-1}(D_1 F^{i+1T}(f^1_{t_{i+1}}, I^1_{t_{i+1}})|\delta f^1_{t_{i+1}}) + (\lambda^{1,f}_{t^-_{i+1}}|\delta f^1_{t^-_{i+1}}) - (\lambda^{1,f}_{t_i}|\delta f^1_{t_i})\end{aligned}$$

where we break the trajectory into each time interval, and perform integration by parts on each interval. This will accommodate the fact that λ may be discontinuous at the times t_i.

Combining these, with $\delta f^0_{1^-} = \delta f^1_{t_1}$, gives at the first timepoint

$$\lambda^{0,f}_{1^-} + D_2 R^{1T}(p^1_{t_1}, f^1_{t_1}) + D_1 F^{1T}(f^1_{t_1}, I^1_{t_1}) - \lambda^{1,f}_{t_1} = 0$$

At other time points

$$D_1 F^{iT}(f^1_{t_i}, I^1_{t_i}) + \lambda^{1,f}_{t^-_i} - \lambda^{1,f}_{t_i} = 0$$

for $i \in \{2, \ldots, N-1\}$. If F^i is independent of the surface $f^1_{t_i}$ (for example, sum of square error of the images) then $\lambda^{1,f}$ is continuous here. Finally at the last timepoint

$$\lambda^{1,f}_{t_N} + D_1 F^{iT}(f^1_{t_N}, I^1_{t_N}) = 0$$

Note that when

$$R^1(p^1_{t_1}, f^1_{t_1} + \epsilon\delta f^1_{t_1}) = \frac{1}{2}\iint_{U\times U} p^{1T}_{t_1}(u)K(f(u), f(u'))p^1_{t_1}(u')dudu' \doteq \frac{1}{2}\|v_0\|^2_V$$

we have

$$\begin{aligned}
&\frac{d}{d\epsilon}R^1(p^1_{t_1}, f^1_{t_1})\Big|_{\epsilon=0}\\
&= \frac{d}{d\epsilon}\frac{1}{2}\iint_{U\times U} p^{1T}_{t_1}(u)K(f(u)+\epsilon\delta f(u), f(u')+\epsilon\delta f(u'))p^1_{t_1}(u')dudu'\Big|_{\epsilon=0}\\
&= \frac{1}{2}\iint_{U\times U} p^{1T}_{t_1}(u)(d_1K(f(u), f(u'))\delta f(u) + d_2K(f(u), f(u'))\delta f(u'))p^1_{t_1}(u')dudu'\\
&= \frac{1}{2}\iint_{U\times U} p^{1T}_{t_1}(u)(d_1K(f(u), f(u'))\delta f(u) + d_1K(f(u'), f(u))\delta f(u'))p^1_{t_1}(u')dudu'\\
&= \iint_{U\times U} p^{1T}_{t_1}(u)p^{1T}_{t_1}(u')d_1K(f(u), f(u'))\delta f(u)dudu'\\
&= (\int_U p^{1T}_{t_1}(u)p^1_{t_1}(u')\nabla_1K(f(u), f(u'))du'|\delta f(u))\\
&\doteq -\frac{d}{dt}p^1_{t_1}
\end{aligned}$$

This is a value we will have already computed to determine the flow, so this equality is important computationally.

9.2.2 Perturbation with respect to I

We consider $I^0 \mapsto I^0 + \epsilon\delta I^0$ with $\delta I^0 = 0$. And we consider $I^1 \mapsto I^1 + \epsilon\delta I^1$. We will require that $\delta I^i_0 = \delta I^0_{1^-}$ (each trajectory is continuous). Note that $\delta I^0_0 = 0$.

For I^0

$$(\lambda^{0,I}_{1^-}|\delta I^0_{1^-}) - (\lambda^{0,I}_0|\delta I^0_0)$$

the second term being zero because we have a fixed template.

And for I^1

$$(D_2F^{1T}(f^1_{t_1}(f^1_{t_1})|\delta I^1_{t_1}) + \sum_{i=1}^{N-1}(D_2F^{i+1T}(f^1_{t_{i+1}}, I^1_{t^-_{i+1}})|\delta I^1_{t^-_{i+1}}) + (\lambda^{1,I}_{t^-_{i+1}}|\delta I^1_{t^-_{i+1}}) - (\lambda^{1,I}_{t_i}|\delta I^1_{t_i})$$

Putting this together with $\delta I^1_{t_1} = \delta I^0_{1^-}$ gives

$$\lambda^{0,I}_{1^-} + D_2F^{1T}(I^1_{t_1}; J_{t_1}) - \lambda^{1,I}_{t_i} = 0$$

Then for the interior points we have

$$\lambda^{1,I}_{t_i} - \lambda^{1,I}_{t_i^-} = D_2 F^{iT}(f^1_{t_i}, I^1_{t_i})$$

for $i \in \{2, \ldots, N-1\}$.

And at the endpoint

$$\lambda^{1,I}_{t_N^-} = -D_2 F^{NT}(f^1_{t_N}, I^1_{t_N})$$

9.2.3 Perturbation with respect to p

We consider $p^0 \mapsto p^0 + \epsilon\delta p^0$ with $\delta p^0 = 0$. And we consider $p^1 \mapsto p^1 + \epsilon\delta p^1$.

For p^0

$$(D_1 R^{0T}(p^0_0, f^0_0)|\delta p^0_0) + (\lambda^{0,p}_{1^-}|\delta p^0_{1^-}) - (\lambda^{0,p}_0|\delta p^0_0)$$

And for p^1

$$(D_1 R^{1T}(p^1_{t_1}, f^1_{t_1})|\delta p^1_{t_1}) + (\lambda^{1,p}_{t_N{}^-}|\delta p^1_{t_N{}^-}) - (\lambda^{0,p}_{t_1}|\delta p^1_{t_1})$$

There is no need to break the trajectory into small intervals here because the fidelity terms will not depend on p. Recall that $\delta p^0_{1^-} \neq \delta p^1_0$.

This implies

$$D_1 R^{0T}(p^0_0, f^0_0) = \lambda^{0,p}_0$$

and

$$\lambda^{0,p}_{1^-} = 0$$

Lastly

$$D_1 R^{1T}(p^1_{t_1}, f^1_{t_1}) = \lambda^{1,p}_{t_1}$$

and

$$\lambda^{1,p}_{t_N^-} = 0$$

9.3 Gradient of the cost function

As before we interpret

$$\frac{d}{dp_1}C = D_1R^{1T}(p^1_{t_1}, f^1_{t_1}) - \lambda^{1,p}_{t_1}$$

so that the gradient is

$$K_{ff}^{-1}(D_1R^{1T}(p^1_{t_1}, f^1_{t_1}) - \lambda^{1,p}_{t_1})$$

and

$$\frac{d}{dp_0}C = D_1R^{0T}(p^0_0, f^0_0) - \lambda^{0,p}_0$$

so that the gradient is

$$K_{ff}^{-1}(D_1R^{0T}(p^0_0, f^0_0) - \lambda^{1,p}_0)$$

9.4 Algorithm for longitudinal image matching with singular geodesic coordinates

This gives the following gradient descent based algorithm for longitudinal image matching. The algorithm can be summarized by first flowing the data forward, second computing cost function gradients, third flowing the Lagrange multipliers backward, and fourth updating the parameters p^0_0, p^1_0. The differences between λ at a time t_i and immediately before it will be represented below by incriminating its value while flowing it backward in time.

Algorithm 9.1. Longitudinal image matching with singular geodesic coordinates

1. Initialize f^0_0 to a template, initialize p^0_0 and $p^1_{t_1}$ (typically to 0)
2. Calculate f^0_t, p^0_t by flowing equations forward in time for $t \in [0, 1)$,
3. calculate f^1_t, p^1_t for $t \in [t_1, t_N]$ being sure to sample each t_i for $i \in \{1, \ldots, N\}$
4. Calculate $v^0_t(x)$, $\varphi^{0,-1}_t(x)$ and $\varphi^{0,-1}_{1t}(x)$ and I^0_t for $t \in [0, 1)$

5. Calculate $v^1_t(x)$, $\varphi^{1,-1}_t(x)$ and I^1_t for $t \in [t_1, t_N]$ being sure to sample each t_i for $i \in \{1, \ldots, N\}$

6. Calculate $\varphi^{1,-1}_{t_{i+1},t}(x)$ for $t \in [t_i, t_{i+1}]$ for $i \in \{1, \ldots, N-1\}$ (for use updating λ^I)

7. Calculate $D_1 F^i(f^1_{t_i}, I^1_{t_i}), D_2 F^i(f^1_{t_i}, I^1_{t_i})$ for $i \in \{1, \ldots, N\}$, calculate $DR^0_1(p^0_0, f^0_0)$, $DR^0_2(p^0_0, f^0_0)$, and $DR^1(p^1_{t_1}, f^1_{t_1})$

8. Initialize $\lambda^{1,f}_{t_N} = -D_1 F^N(f^1_{t_N}, I^1_{t_N}), \lambda^{1,I}_{t_N} = -D_2 F^N(f^1_{t_N}, I^1_{t_N}), \lambda^{1,p}_{t_N} = 0$

9. For $i \in \{N-1, \ldots, 1\}$

    (a) Calculate $\lambda^{1,I}_t$ explicitly for $t \in [t_i, t_{i+1}]$

    (b) Flow $\lambda^{1,f}_t, \lambda^{1,p}_t$ backwards from t_{i+1} to t_i

    (c) Subtract $D_1 F^{iT}(f^1_{t_i}, I^1_{t_i};)$ from $\lambda^{1,f}_{t_i}$ and $D_2 F^{iT}(f^1_{t_i}, I^1_{t_i};)$ from $\lambda^{1,I}_{t_i}$

10. Calculate gradient vector $K^{-1}_{f^1_0 f^1_0}(D_1 R^1(p^1_{t_1}, f^1_{t_1}) - \lambda^{1,p}_{t_1})$

11. Set $\lambda^{0,f}_1$ to $\lambda^{1,f}_{t_1} - D_2 R^{1T}(p^1_{t_1}, f^1_{t_1})$, set $\lambda^{0,p}_1$ to 0, set $\lambda^{0,I}_1 = \lambda^{1,I}_{t_1}$ (noting we have already performed the subtraction in 9c)

12. Calculate $\lambda^{0,I}_t$ explicitly for $t \in [0, 1]$

13. Flow $\lambda^{0,f}_t, \lambda^{0,p}_t$ backwards from $t = 1$ to $t = 0$

14. Calculate gradient vector $K^{-1}_{f^0_0 f^0_0}(D_1 R^{0T}(p_0) - \lambda^p_0)$

15. Update $p^0_0 \mapsto p^0_0 + \epsilon K^{-1}_{f^0_0 f^0_0}(D_1 R^{0T}(p^0_0, f^0_0) - \lambda^{0,p}_0)$ for some small ϵ

16. Update $p^1_{t_1} \mapsto p^1_{t_1} + \epsilon K^{-1}_{f^1_0 f^1_0}(D_1 R^{1T}(p^1_{t_1}, f^1_{t_1}) - \lambda^{1,p}_{t_1})$ for some small ϵ

17. Repeat steps 2 to 2 until convergence criteria is met.

Note in step 6 that since we have to increment λ^I at each t_i, we will deform the image from one t_i to the previous, rather than from the endpoint t_N to each.

9.5 Examples

We present the results of algorithm 8.1 and 9.1 applied to a volumetric study of the entorhinal cortex and trans entorhinal cortex in the ADNI population. This work was presented at [109], to be published in Lecture Notes in Computer Science, and much of the content below appears there.

9.5.1 Reducing variability in volumetric analysis of the entorhinal and trans entorhinal cortex in the ADNI dataset

Data

T1 brain MR images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset were examined[1]. Twenty patients were selected, older adults (age 72 ± 8 years), 60% male, education of 17 ± 3 years, with mild cognitive impairment, and having a continuous left collateral sulcus. Each subject was scanned up to 5 times, and at least 3 times (so that a residual can be estimated after linear fitting), over 2 years. The entorhinal cortex and immediately lateral cortex (the trans entorhinal cortex) were analyzed for the presence of atrophy. Structures were delineated by manual segmentation, as well as by FreeSurfer for comparison with state of the art [33, 104].

Imaging data for each subject was rigidly aligned to baseline by minimizing sum of square error in T1 images, and imaging data between subjects was rigidly aligned to a single subject through 4 landmarks placed automatically at the boundaries of the segmentations. For each subject i, at time t_j, we denote the rigidly aligned manual segmentation image as J^{ij}.

The template used in this analysis was estimated as described in Sec. 8.4.1

---

[1] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see `www.adni-info.org`.

We use the following log-linear model to estimate volumetric atrophy rate in each subject

$$\log v^{ij} = a_0^i + a_1^i t^{ij} + \epsilon^{ij}$$

where v^{ij} is subject i's entorhinal cortex and trans entorhinal cortex volume at time t^{ij}, a_0^i is a nuisance parameter (log volume at $t = 0$), and a_1^i is the atrophy rate (exponential time constant). In this model, ϵ^{ij} is assumed to be independent Gaussian noise with variance σ_i^2.

Each parameter is estimated by maximum likelihood, including σ_i^2 which is the mean square error of the fit. The variance of our atrophy rate estimator is given by

$$\mathrm{Var}[\hat{a}_1^i] = \frac{\hat{\sigma}_i^2}{N_{ti}\sigma_{ti}^2} \tag{9.1}$$

where N_{ti} is the number of timepoints for subject i, σ_{ti}^2 is the variance in scan times, and $\hat{\sigma}_i^2$ is our estimate of the variance in ϵ^{ij}. Note that (9.1) agrees with the residual bootstrap variance estimator within 3.3% (root mean square percent error). We compare the standard deviation of this estimator measured from manual segmentations, to that measured after our filtering procedure, and to that computed from FreeSurfer.

Note that because we are performing estimation in each subject, we do not require a random effects model.

Mapping results

The estimated entorhinal cortex and trans-entorhinal cortex atlas is shown in cyan in Fig. 9.1. Two example longitudinal maps are shown in Fig. 9.1, illustrating a high variability case (top), and a low variability case (bottom). Note the difference in anterior-posterior (left-right on the figure) extent in the manual segmentations (red) for the first two timepoints for the high variability subject. This inconsistency has been filtered out by our mapping procedure (blue).

Atrophy rate

For the two subjects shown in Fig. 9.1, volumetric analysis is shown in Fig. 9.2. Volumes of the manual segmentations are shown as red dots, while volumes of the deforming template are shown

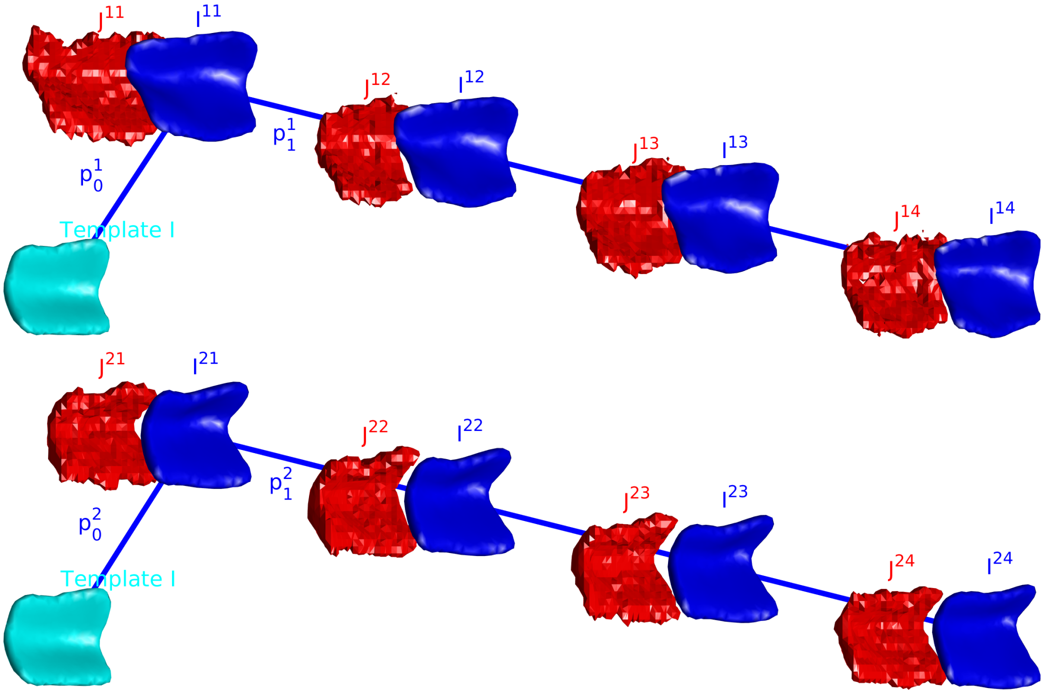


Figure 9.1: Example longitudinal mapping results for two subjects. The template is shown in cyan, the deforming template in blue, and the target in red. Top: high variability example, bottom: low variability example.

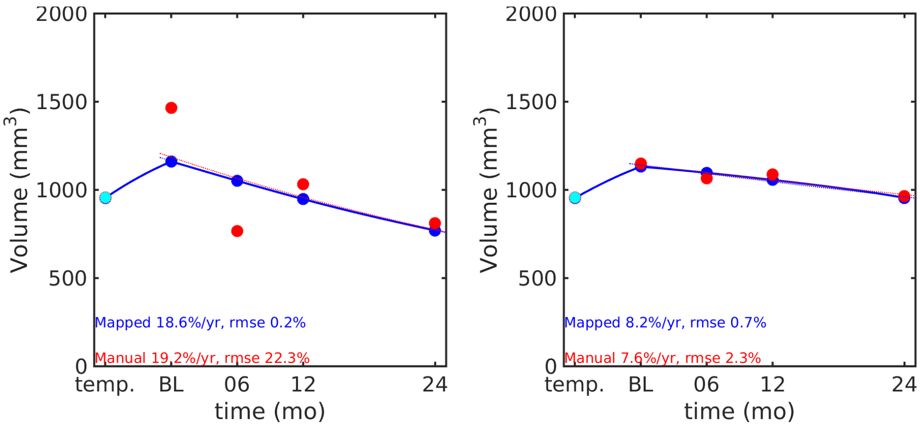


Figure 9.2: The mapping procedure's stabilization of volume measurements is illustrated for left: high variability example, and right: low variability example.

as a blue line. The volume of the deformed template corresponding to each measured timepoint is shown as a blue dot, and the volume of the template itself is shown on the left. The reduction in variance due to the longitudinal mapping procedure is evident, particularly for the highly variable subject (right).

The atrophy rate estimated for each subject is shown in Fig. 9.3, with manual segmentations shown in red, the results of our longitudinal mapping procedure shown in blue, and results from FreeSurfer shown in green for comparison with state of the art.

Atrophy rate estimators are quite consistent between the manual segmentations and the longitudinal maps (correlation coefficient 0.9881), and not very consistent with FreeSurfer results (correlation coefficient 0.2283), as can be seen in the scatter plot in Fig. 9.4.

The standard deviation of our atrophy rate estimator, computed according to (9.1), is shown in Fig. 9.5. Significant differences between the three methods are determined by pairwise signed rank tests. Variance is significantly reduced in longitudinal maps relative to manual segmentations ($p =$ 8.86e-05) and relative to FreeSurfer ($p =$ 1.03e-04). However, variance in FreeSurfer estimates is not significantly different from manual segmentations ($p =$6.81e-01).

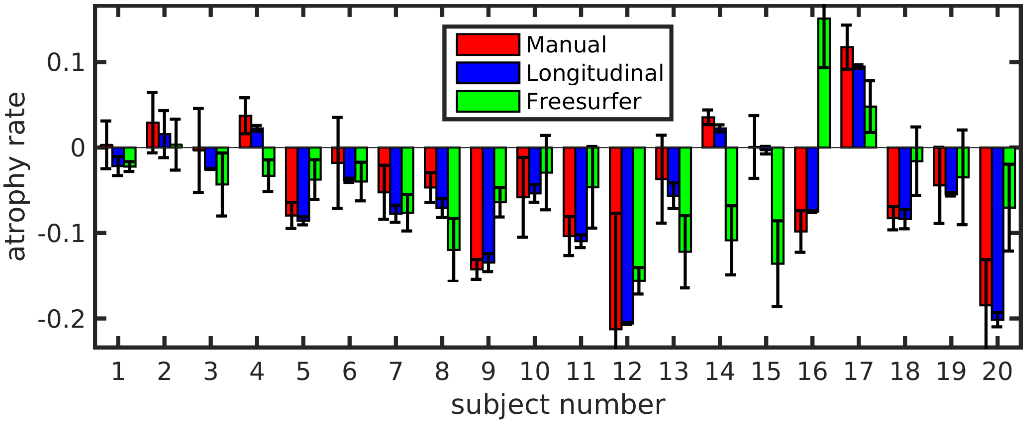


Figure 9.3: Estimated atrophy rate is shown for each subject examined as a bar, while standard deviation of the estimator is shown as an errorbar.

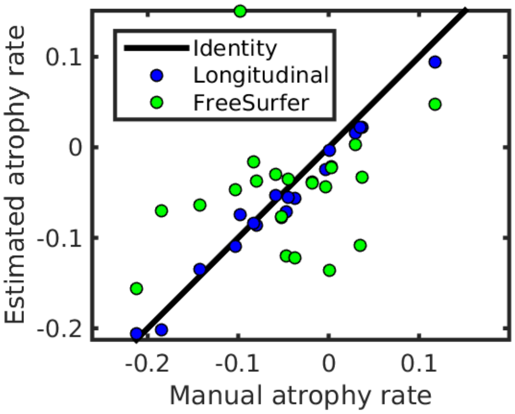


Figure 9.4: Correlation between atrophy rate measured from manual segmentations (x axis) and estimated with two methods (y axis) is visualized with a scatter plot. Correlation coefficient for longitudinal maps: 0.9881, and for FreeSurfer: 0.2283.

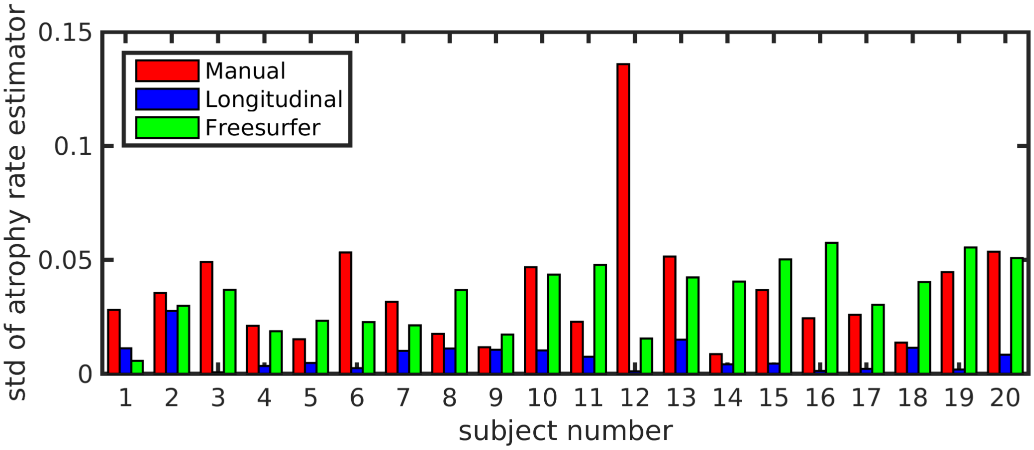


Figure 9.5: Standard deviation of our atrophy rate estimator, for each of the 20 subjects examined, is shown for volume of manual segmentations (red) and volume of our template deformed by the longitudinal mapping procedure.

Discussion

As our population of manual segmentations expands to include healthy control subjects in addition to those with MCI, we intend to employ this procedure to identify changes that are specific to disease, as opposed to normal aging. Local modeling of tissue change based on determinant of Jacobian of our mappings will likely prove more sensitive than the volumetry presented here, and can be expanded to include volume change (determinant of 3×3 Jacobian), surface area change (determinant of the 2×2 component of the Jacobian tangent to the template surface), and thickness change (determinant of the 1×1 component of the Jacobian normal to the template surface). An example of early work with this type of analysis can be found at [110].

Because our method treats each segmentation differently depending on its position in the time-series, unlike the approach in longitudinal FreeSurfer as mentioned in the introduction, the potential for processing bias exists. This was estimated by reversing the order of the timeseries and repeating the experiment, showing an average overestimate in the magnitude of atrophy rate constant of 0.01. This is a small source of error relative to the inconsistencies in anatomical definitions over time we have sought to address. One simple approach for removing this source of bias is to take the average of the forwards and reversed atrophy rates. A second would be to choose the one which produces the smaller value of the cost function. These and other strategies for removing bias will be the subject of future research.

The longitudinal mapping procedure presented here is able to filter segmentation images, significantly reducing uncertainty in atrophy rate measurements, while correlating strongly with raw manual segmentation results. This procedure has important implications for clinical studies of Alzheimer's disease, where reduced variability will allow for sufficient statistical power at smaller sample sizes.

9.5.2 Diffeomorphometry of the entorhinal and trans entorhinal cortex

Using the same dataset described above, we model atrophy at the population level. We use the same log linear model at each vertex, and add a random effect corresponding to the patient identity. This divides the variability into inter-patient (the random effect), and inter-scan (independent noise between observations).

We visualize atrophy rate as a percent per year as $100(1 - \exp(a_1))$.

Volume atrophy

Volume atrophy, or determinant of 3×3 Jacobian, was used as a biomarker for shape change. This is shown in Fig. 9.6

Area atrophy

Area atrophy, or determinant of 2×2 tangential Jacobian, estimated from change in triangle area in template triangulated surface, interpolated from faces to vertices by adding 1/3 of each neighboring face (such that the sum over vertices is equal to the sum over faces) was used as a biomarker for shape change. This is shown in Fig. 9.7.

Thickness atrophy

Thickness atrophy, or determinant of 1×1 normal Jacobian, estimated as the ratio of the two previous quantities, was used as a biomarker for shape change. This is shown in Fig. 9.8.

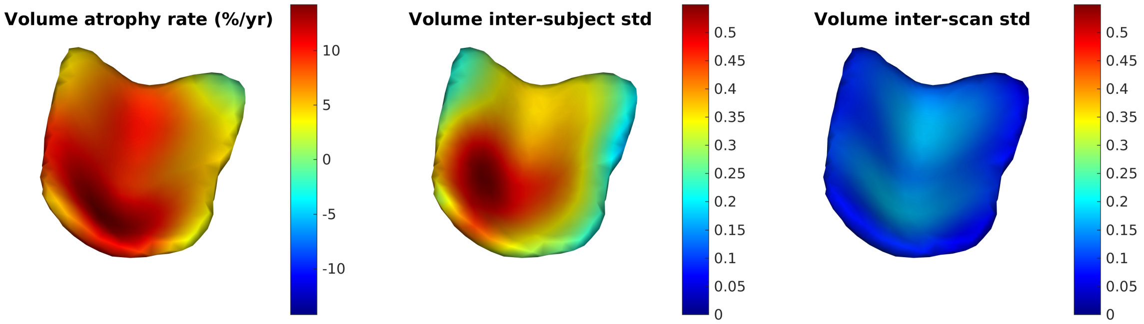


Figure 9.6: Volume atrophy rate in the entorhinal cortex and trans entorhinal cortex are shown on the left. The two components of its variance, inter-subject and inter-scan, are shown in the middle and right respectively on the same color scale. Left: lateral, right: medial, up: rostral, down: caudal, view from superior.

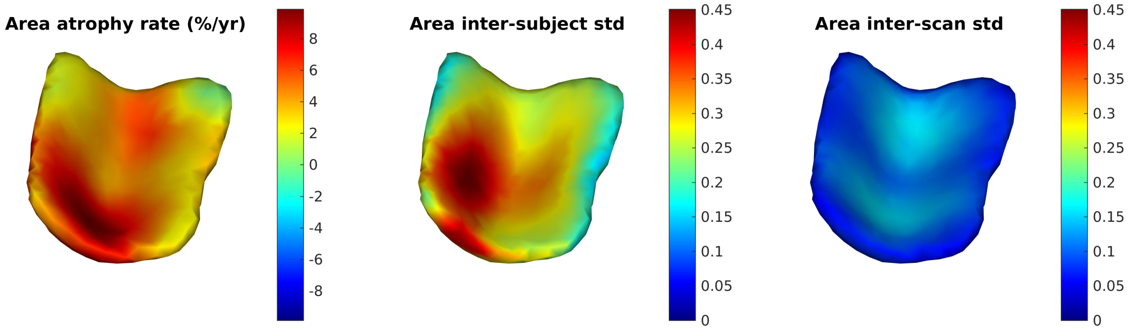


Figure 9.7: Area atrophy rate in the entorhinal cortex and trans entorhinal cortex are shown on the left. The two components of its variance, inter-subject and inter-scan, are shown in the middle and right respectively on the same color scale. Left: lateral, right: medial, up: rostral, down: caudal, view from superior.

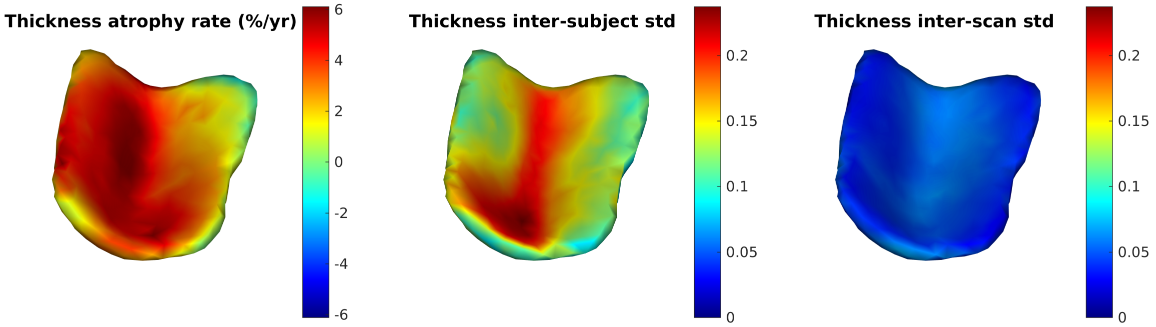


Figure 9.8: Thickness atrophy rate in the entorhinal cortex and trans entorhinal cortex are shown on the left. The two components of its variance, inter-subject and inter-scan, are shown in the middle and right respectively on the same color scale. Left: lateral, right: medial, up: rostral, down: caudal, view from superior.

Notice most thickness atrophy is found caudally, as well as laterally in the sulcal ERC and in the TEC.

Discussion

This data shows the spatial distribution of atrophy in the entorhinal region in subjects with mild cognitive impairment. Regions with the highest amount of atrophy are lateral, consistent with pathology reported by Braak and Braak [3]. They are also more caudal, a pattern currently being explored.

When manual segmentations of control subjects has been completed, future work will include hypothesis testing at each vertex, determining in which areas MCI subjects are experiencing atrophy at a rate different from controls.

Chapter 10

General Discussion

In this work we developed an efficient representation for diffeomorphic mappings, with information concentrated on bounding surfaces of relevant anatomical structures, and built tools for overcoming challenges associated to common datasets of dense images: noisy or anomalous data, images with multiple segmentation labels, and timeseries of images.

This approach is in line with two broad trends in data science: exploiting sparsity, and working with big data. Sparse models for high dimensional data such as medical images have the advantage of avoiding the curse of dimensionality, allowing estimation and hypothesis testing to be performed with realistically sized datasets. In our approach to reducing dense 3D deformations to 2D surface representations, we build sparsity through an understanding of the interplay between flows of diffeomorphisms and properties of data used in neuroimaging. The sparse model is optimal in many scenarios, and a good approximation in others, because anatomy at the millimeter scale can be well described by structures with homogeneous intensities and smooth boundaries. Techniques such as ℓ_1 minimization, which provide sparse outputs without an explanation of “why” are not required. The surface based diffeomorphometry presented here is sufficient to detect subtle differences between populations, and is easily interpreted and visualized. Alternative examples of sparse representations can be found in the literature, including a bandlimited representation of velocity fields by Zhang, Wells III, and Golland [111], and a control points approach by Durrleman, Allassonnière, and Joshi

[90] discussed in Section 5.7.5.

When discussing big data one often glosses over the fact that what is needed is "good big data". In published neuroimage datasets, large numbers of images with multiple structures segmented often require a prohibitive amount of quality control before they become "good". Many examples shown in this work illustrate the necessary trade-offs between "big" and "good". The techniques developed here allow us to work with data that has not been quality controlled, by filtering out artifacts that are unnatural. For example, segmentations in the PREDICT study had too many outliers that needed to be excluded from analysis to draw meaningful conclusions. Our methods allowed Faria et al. [112] to use the entire dataset, and draw important conclusions about the progression of Huntington's disease. A trend in studying anatomy with diffeomorphisms has been moving toward modeling of timeseries, as discussed in Section 9. Continuing with this theme, our contribution to the area has been in approaching timeseries modeling as an opportunity for filtering inconsistencies in highly variable data.

One important issue not discussed in this thesis is computation time. The compact representation of diffeomorphisms used here does not translate to faster computation. Much of the computation is performed in dense 3D space (e.g. calculating φ^{-1} from v), or interpolating between sparse surfaces and dense space (e.g. calculating v from p), and is more time consuming than corresponding fully dense algorithms which can benefit from fast Fourier transforms to speed some calculations. To address this, a substantial amount of effort has been put toward quantifying computational performance, and developing parallelization schemes to benefit from GPUs and high performance computing resources. Examples of our work in this area are [113–117]. As image resolution improves, populations sizes increase, and the complexity of anatomical atlases grows, the issues of computation time requires continued attention.

Another limitation is that we have focused almost exclusively on labeled images, putting our efforts into modeling φ and not into modeling the J term in the random orbit model. There are many approaches to developing this term, including physics based models of imaging processes employed by FreeSurfer. Many other techniques for dealing with heterogeneity of imaging data are model-free, involving histogram matching or local histogram matching, fidelity terms that are invariant to linear intensity transformations such as used in AIR, or based on mutual information [118], or neighborhood based self similarity [119, 120]. These approaches become more important when dealing with mapping between images from different modalities.

The notion of inter modality registration will be important for our future work. An exciting direction in neuroimaging research involves integrating data at multiple scales. This includes 3D histology at the micron scale for visualizing individual neurons and pathology, high resolution ex vivo imaging at the 100 micron scale for visualizing specific nuclei and cortical subfields, and population level MRI at the mm scale for visualizing gross anatomy. We intend to apply the methods developed in this thesis to unite each of these modalities in a common coordinate system, providing direct evidence of the relationship between morphometric imaging biomarkers and pathological processes.

While curing neurodegenerative disease is beyond the hope of image analysts, an important step along the way will be to use structural imaging to quantify the success of a clinical intervention. Our techniques will be used in a clinical trial for an early Alzheimer's intervention based on the work described in 7.3.1 and [94]. At the population level, structural imaging biomarkers may be sensitive enough to demonstrate efficacy of the intervention before cognitive changes are measurable. This has the potential to reduce the duration, and therefore cost, of such a study.

Appendix A

Solution to Beg's image matching problem

We will show the necessary conditions for solving Prob. 1. This was not the approach originally used to solve the problem, but it is consistent with what will be used elsewhere in the thesis.

We let the state of our system be determined by the function $I_t : \Omega \in \mathbb{R}^3 \to \mathbb{R}^C$ for an image with C channels. With $I_t = I_0 \circ \varphi_t^{-1}$, we can write the dynamics for I_t using the optical flow equation (C.4) $\frac{d}{dt} I_t = -DI_t v_t$.

We will enforce these dynamics with Lagrange multipliers, $\lambda_I(t)$, and find a stationary point to the augmented cost function

$$\int_0^1 \frac{1}{2\sigma_R^2} \|v_t\|_V^2 dt + \frac{1}{2\sigma_M^2} \|I_1 - J\|_{L_2}^2 + \int_0^1 \int_\Omega \lambda_t^{IT}(x) [\frac{d}{dt} I_t(x) + DI_t(x) v_t(x)] dx dt$$

Statement 11 (Necessary conditions for solving Beg's image matching problem)**.** Stationary solutions to the image matching problem are characterized by

$$\frac{1}{\sigma_R^2}(L^* L) v_t(x) - \frac{1}{\sigma_M^2} DI_t^T(x)(I_0 \circ \varphi_t^{-1} - J \circ \varphi_{1t}^{-1}) |D\varphi_{1t}^{-1}| = 0$$

Proof. We will first take a perturbation $I_t \mapsto I_t + \epsilon \delta I_t$ with $\delta I_0 = 0$ (fixed template), and show conditions for a stationary solution

$$\begin{aligned}&\frac{d}{d\epsilon}\frac{1}{2\sigma_M^2}\|I_1+\epsilon\delta I_1-J\|^2_{L_2}+\int_0^1\int_\Omega \lambda_t^{IT}(x)[\frac{d}{dt}I_t(x)+\epsilon\frac{d}{dt}\delta I_t(x)+DI_t(x)v_t(x)+\epsilon D\delta I_t(x)v_t(x)]dxdt\Big|_{\epsilon=0}\\
&=\frac{1}{\sigma_M^2}\int_\Omega(I_1(x)-J(x))^T\delta I_1(x)dx+\int_0^1\int_\Omega \lambda_t^{IT}(x)\frac{d}{dt}\delta I_t(x)dxdt+\int_0^1\int_\Omega \lambda_t^{IT}(x)D\delta I_t(x)v_t(x)]dxdt\\
&=\frac{1}{\sigma_M^2}\int_\Omega(I_1(x)-J(x))^T\delta I_1(x)dx-\int_0^1\int_\Omega\frac{d}{dt}\lambda_t^{IT}(x)\delta I_t(x)dxdt+\int_\Omega\lambda_1^{IT}\delta I_1dx-\int_0^1\int_\Omega\mathrm{div}[v(x)\lambda_t^{IT}(x)]\delta I_t(x)]dxdt\end{aligned}$$

where apply integration by parts and the divergence theorem, assuming zero boundary conditions in the second case.

For this to be stationary with respect to any perturbation we will require boundary conditions

$$\lambda_1^I+\frac{1}{\sigma_M^2}(I_1-J)=0$$

as well as dynamics

$$\frac{d}{dt}\lambda_t^I+\mathrm{div}[v_t\lambda_t^{IT}]^T=0$$

As shown in C.4 this equation has an explicit solution given the diffeomorphism, $\lambda_t^I=\lambda_1^I(\varphi_t^{-1})|D\varphi_t^{-1}|$.

In our problem we have boundary conditions specified at $t=1$, so we will use the form

$$\lambda_t^I=\lambda_1^I(\varphi_{1t}^{-1})|D\varphi_{1t}^{-1}|$$

where $\varphi_{1t}=\varphi_t(\varphi_1^{-1})$ (i.e. from time 1 back to time zero, and then from time 0 to time t). With this in mind we can write $\lambda_t=-\frac{1}{\sigma_M^2}(I_1\circ\varphi_{1t}^{-1}-J\circ\varphi_{1t}^{-1})|\varphi_{1t}^{-1}|=-\frac{1}{\sigma_M^2}(I_0\circ\varphi_t^{-1}-J\circ\varphi_{1t}^{-1})|\varphi_{1t}^{-1}|$.

Next we take a perturbation $v\mapsto v+\epsilon\delta v$. This gives

$$\begin{aligned}&\frac{d}{d\epsilon}\frac{1}{2\sigma_R^2}\int_0^1\int_\Omega[(L^*L)v_t(x)+\epsilon(L^*L)\delta v_t(x)]^T(v_t(x)+e\delta v_t(x))dxdt+\int_0^1\int_\Omega\lambda_t^{IT}(x)[DI_t(x)v_t(x)+\epsilon DI_t(x)\delta v_t(x)]dxdt\Big|\\
&=\int_0^1\int_\Omega\frac{1}{\sigma_R^2}[(L^*L)v_t(x)]^T\delta v_t(x)dxdt+\int_0^1\int_\Omega\lambda_t^{IT}(x)DI_t(x)\delta v_t(x)dxdt\end{aligned}$$

For a stationary solution we require

$$\frac{1}{\sigma_R^2}(L^*L)v_t(x)+DI_t^T(x)\lambda_t^I=0$$

Combining with the previous result gives

$$\frac{1}{\sigma_R^2}(L^*L)v_t(x) - \frac{1}{\sigma_M^2}DI_t^T(x)(I_0 \circ \varphi_t^{-1} - J \circ \varphi_{1t}^{-1})|D\varphi_{1t}^{-1}| = 0$$

This is the variation of the cost function with respect to a perturbation in v_t, obeying the dynamical constraints linking v_t and I_t. This variation is used to build a gradient descent algorithm in Beg et al. [21].

Notice that the quantity $(L^*L)v_t$ is proportional to the gradient of the deforming image at all times, with the proportionality being an error term. In some sense v_t “goes with the flow”. This observation motivates the use of a more compact description of φ, in terms of the initial velocity v_0 only. This discussion is continued in Appendix B.

Appendix B

Initial momentum representation

We quantify this behavior by noting that any solution of Prob. 1 involves a length minimizing trajectory or geodesic. That is, for any fixed value of φ_1, we can lower the cost of our solution by making the trajectory a constant speed geodesic. In other words we wish to minimize

$$\frac{1}{2}\int_0^1 \|v_t\|_V^2 dt$$

over φ_t, v_t with the dynamical constraint $\frac{d}{dt}\varphi_t = v_t(\varphi_t)$, and the boundary constraints φ_0 and φ_1 fixed. We introduce the augmented cost function

$$\int \frac{1}{2}\|v_t\|_V^2 dt + \int_0^1 \int_\Omega \lambda_t^{\varphi T}(x)[\frac{d}{dt}\varphi_t(x) - v_t(\varphi_t(x))]dxdt$$

Statement 12 (Necessary conditions for geodesics)**.** Geodesic trajectories are described in terms of Lagrangian momentum λ^φ

$$(L^*L)v = \lambda^\varphi(\varphi^{-1})|D\varphi^{-1}|$$

$$\frac{d}{dt}\lambda^\varphi = -Dv^T(\varphi)\lambda^\varphi$$

They are described in terms of the Eulerian momentum m by

$$(\text{Ł}^*L)v = m$$

$$\frac{d}{dt}m = -(Dmv + m\,\mathrm{div}\,v + Dv^T m)$$

Specifying λ_0^φ or m_0 defines the entire trajectory.

Proof. First we take a perturbation with respect to v_t.

$$
\begin{aligned}
&\frac{d}{d\epsilon}\frac{1}{2}\int_0^1\int_\Omega [(L^*L)v_t(x)+\epsilon(L^*L)\delta v_t(x)]^T(v_t(x)+\epsilon\delta v_t(x))dxdt - \int_0^1\int_\Omega \lambda_t^{\varphi T}(x)(v_t(\varphi_t)+\epsilon\delta v_t(\varphi_t))dxdt\bigg|_{\epsilon=0}\\
&=\int_0^1\int_\Omega [(L^*L)v_t(x)]^T\delta v_t(x)dxdt - \int_0^1\int_\Omega \lambda_t^{\varphi T}(x)\delta v_t(\varphi_t(x))dxdt\\
&=\int_0^1\int_\Omega [(L^*L)v_t(x)]^T\delta v_t(x)dxdt - \int_0^1\int_\Omega \lambda_t^{\varphi T}(\varphi_t^{-1}(y))\delta v_t(y)|D\varphi^{-1}(y)|dydt
\end{aligned}
$$

where we let $y=\varphi_t(x)$ so that $x=\varphi_t^{-1}(y)$ and $dx=|D\varphi_t^{-1}(y)|$. For this to be stationary we must have

$$
(L^*L)v_t(x)-\lambda_t^{\varphi}(\varphi_t^{-1}(x))|D\varphi_t^{-1}(x)|=0 \tag{B.1}
$$

This gives an important relationship between v and λ^φ. We can also write

$$
\begin{aligned}
\lambda_t^\varphi(x) &= (L^*L)v_t(\varphi_t(x))/|D\varphi_t^{-1}(\varphi_t(x))|\\
&= (L^*L)v_t(\varphi_t(x))|[D\varphi_t^{-1}(\varphi_t(x))]^{-1}|\\
&= (L^*L)v_t(\varphi_t(x))|D\varphi_t(x)|
\end{aligned}
$$

where the last line follows from C.1. The quantity λ_t^φ is often called the Lagrangian momentum, since it is indexed to the deforming coordinate system through φ_t.

Now taking a perturbation with respect to φ gives

$$
\begin{aligned}
&\frac{d}{d\epsilon}\int_0^1\int_\Omega \lambda_t^{\varphi T}(x)[\frac{d}{dt}\varphi_t(x)+\epsilon\frac{d}{dt}\delta\varphi_t(x)-v_t(\varphi_t(x)+\epsilon\delta\varphi_t(x))]dxdt\bigg|_{\epsilon=0}\\
&=\int_0^1\int_\Omega \lambda_t^{\varphi T}(x)[\frac{d}{dt}\delta\varphi_t(x)-Dv_t(\varphi_t(x))\delta\varphi_t(x)]dxdt\\
&=-\int_0^1\int_\Omega \frac{d}{dt}\lambda_t^{\varphi T}(x)\delta\varphi_t(x)dxdt-\int_0^1\int_\Omega \lambda_t^{\varphi T}(x)Dv_t(\varphi(x))\delta\varphi_t(x)dxdt
\end{aligned}
$$

Note that there are no boundary conditions to include in the integration by parts because $\delta\varphi_0=\delta\varphi_1=0$ (fixed endpoints). For stationarity, this condition implies that

$$
\frac{d}{dt}\lambda_t^\varphi(x)+Dv_t^T(\varphi_t(x))\lambda_t^\varphi(x)=0 \tag{B.2}
$$

This is a simple evolution equation for the Lagrangian momentum, $\frac{d}{dt}\lambda^\varphi=-Dv^T(\varphi)\lambda^\varphi$. For an initial condition λ_0^φ, we can solve for λ_t^φ and v_t, from which we can recover φ_t.

The equation (B.2) is often rewritten in a Eulerian form in terms of v. The first term becomes

$$\begin{aligned}
&\frac{d}{dt}[(L^*L)v_t(\varphi_t(x))|D\varphi_t(x)|] \\
&= \frac{d}{dt}[(L^*L)v_t(\varphi_t(x))]|D\varphi_t(x)| + (L^*L)v_t(\varphi_t(x))\frac{d}{dt}|D\varphi_t(x)| \\
&= \frac{d}{dt}(L^*L)v_t(\varphi_t(x))|D\varphi_t(x)| + D(L^*L)v_t(\varphi_t(x))\frac{d}{dt}\varphi_t(x)|D\varphi_t(x)| + (L^*L)v_t(\varphi_t(x))\operatorname{div} v_t(\varphi_t(x))|D\varphi_t(x)| \\
&= \frac{d}{dt}(L^*L)v_t(\varphi_t(x))|D\varphi_t(x)| + D(L^*L)v_t(\varphi_t(x))v_t(\varphi_t(x))|D\varphi_t(x)| + (L^*L)v_t(\varphi_t(x))\operatorname{div} v_t(\varphi_t(x))|D\varphi_t(x)|
\end{aligned}$$

where the equality in the third term involves the relationship between divergence and determinant of Jacobian as shown in C.5, and the equality in the second term is simply the dynamics $\frac{d}{dt}\varphi = v(\varphi)$.

And since the second term in (B.2) can be written as

$$Dv_t^T(\varphi_t(x))(L^*L)v_t(\varphi_t(x))|D\varphi_t(x)|$$

we can cancel $|D\varphi_t(x)|$ from each term, and compose with φ_t^{-1} to give the equation

$$\frac{d}{dt}(L^*L)v_t + [D(L^*L)v_t]v_t + (L^*L)v_t \operatorname{div} v_t + Dv_t^T(L^*L)v_t = 0$$

The quantity $(L^*L)v_t$ is often called the Eulerian momentum m_t, and its geodesic equation can be written as

$$\frac{d}{dt}m + Dmv + m\operatorname{div} v + Dv^T m = 0$$

In this case we can define the initial condition mu_0, and calculate m_t and v_t, and then integrate to give φ_t. However, as shown in the solution to Beg's problem, m_t is proportional to DI_t^T. When I is a binary segmentation of an anatomical structure, m is singular and supported on the structure's boundary, and Dm is difficult to define. For this reason, we adopt a model for geodesic flows related to singular momentum representation. This reduces complexity of our representation further. First from a time varying vector field to an initial vector field. And now from a dense vector field to a sparse representation.

Appendix C

Useful formulas

C.1 Spatial derivative of inverse

Here we derive relationships between the spatial derivative of a diffeomorphism and its inverse.

On the one hand

$$D[\varphi(\varphi^{-1})] = D\,\mathrm{Id} = \mathbf{I} \tag{C.1}$$

where Id is the identity map $\mathrm{Id}(x) \mapsto x$ and $\mathbf{I}$ is the identity matrix. On the other hand

$$D[\varphi(\varphi^{-1})] = D\varphi(\varphi^{-1}) D\varphi^{-1} \tag{C.2}$$

Equating (C.1) and (C.2) gives

$$D\varphi^{-1} = [D\varphi(\varphi^{-1})]^{-1}$$

Alternatively, if I had composed in the other order

$$D[\varphi^{-1}(\varphi)] = D\varphi^{-1}(\varphi) D\varphi \tag{C.3}$$

which gives when combining (C.1) and (C.3)

$$D\varphi^{-1}(\varphi) = [D\varphi]^{-1}$$

These two results are generally the most useful forms, but we can also swap φ with a dummy φ^{-1}, giving

$$D\varphi = [D\varphi^{-1}(\varphi)]^{-1}$$

and

$$D\varphi(\varphi^{-1}) = [D\varphi^{-1}]^{-1}$$

C.2 Time derivative of inverse

Here we show the time derivative of φ_t^{-1}. We begin with the identity,

$$\frac{d}{dt}[\varphi_t(\varphi_t^{-1})] = \frac{d}{dt}\,\mathrm{Id} = 0$$

On the other hand, using $\frac{d}{dt}\varphi_t = v_t(\varphi_t)$

$$\begin{aligned}\frac{d}{dt}[\varphi_t(\varphi_t^{-1})] &= \frac{d}{dt}\varphi_t(\varphi_t^{-1}) + D\varphi_t(\varphi_t^{-1})\frac{d}{dt}\varphi_t^{-1}\\ &= v_t(\varphi(\varphi^{-1})) + D\varphi_t(\varphi_t^{-1})\frac{d}{dt}\varphi_t^{-1}\\ &= v_t + [D\varphi^{-1}]^{-1}\frac{d}{dt}\varphi_t^{-1}\end{aligned}$$

where we used the formula in C.1.

Equating the two gives

$$\frac{d}{dt}\varphi_t^{-1} = -D\varphi_t^{-1}v_t$$

C.3 Optical flow equation

Here we derive how the optical flow equations arises when time varying diffeomorphisms are applied to images.

The image I_t is deformed by the diffeomorphism φ_t

$$I_t = I_0 \circ \varphi_t^{-1}$$

so that it can be described by a dynamical equation simply using the chain rule

$$\frac{d}{dt}I_t = DI_0 \circ \varphi_t^{-1} \frac{d}{dt}\varphi_t^{-1}$$

With $\frac{d}{dt}\phi_t = v_t(\phi_t)$, we show in Sec. C.2 that $\frac{d}{dt}\varphi_t^{-1} = -D\varphi_t^{-1} v_t$, which implies

$$\begin{aligned}\frac{d}{dt}I_t &= -DI_0 \circ \varphi_t^{-1} D\varphi_t^{-1} v_t \\ &= -D[I_0 \circ \varphi^{-1}]v_t \\ &= -DI_t v_t \end{aligned} \tag{C.4}$$

C.4 Adjoint to optical flow

We show that the equation $\frac{d}{dt}\lambda_t^I = -\operatorname{div}[v_t \lambda_t^{IT}]^T$ has an explicit solution in terms of the associated diffeomorphism φ_t.

First note that for an image I_t the quantity $\int \lambda_t^{IT}(x) I_t(x) dx$ is conserved over time. To see this

$$\begin{aligned}\frac{d}{dt}\int_\Omega \lambda_t^{IT}(x) I_t(x) dx &= \int \frac{d}{dt}\lambda_t^{IT} I_t(x) dx + \int \lambda_t^{IT} \frac{d}{dt} I_t dx \\ &= \int_\Omega -\operatorname{div}[\lambda_t^I v_t]^T I_t dx + \int -\lambda_t^{IT} DI_t v_t dx \\ &= \int_\Omega \lambda_t^{IT} DI_t v_t dx + \int -\lambda_t^{IT} DI_t v_t dx \\ &= 0\end{aligned}$$

since $-\operatorname{div}$ is adjoint to D (assuming the normal component of v is zero on the boundary of Ω).

Therefore

$$\begin{aligned}\int_\Omega \lambda_0^{IT} I_0 dx &= \int_\Omega \lambda_t^{IT} I_t dx \\ &= \int_\Omega \lambda_t^{IT}(x) I_0(\varphi_t^{-1}(x)) dx \\ &= \int_\Omega \lambda_t^{IT}(\varphi_t(y)) I_0(y) |D\varphi_t(y)| dy\end{aligned}$$

For this to be true for any image I we must have $\lambda_0^I = \lambda_t^I(\varphi_t)|D\varphi_t|$ so that

$$\lambda_t^I = \lambda_0^I(\varphi_t^{-1}) / |D\varphi_t(\varphi_t^{-1})|$$

$$= \lambda_0^I(\varphi_t^{-1})|[D\varphi_t(\varphi_t^{-1})]^{-1}|$$

$$= \lambda_0^I(\varphi_t^{-1})|D\varphi_t^{-1}|$$

using C.1.

C.5 Divergence and determinant

The divergence, $\operatorname{div} v$, tells us how expansive a vector field is, and the determinant of Jacobian, $|D\varphi|$, tells us how expansive a diffeomorphism is. We seek a relationship between the two. We will look at the quantity $\int w(\varphi^{-1})(x)|D\varphi(x)|dx$ and ask how this changes in time.

We start with a smooth scalar test function w_0. We define $w_t = w_0(\varphi_t^{-1})$. We want to know how its integral changes over time (or how its volume changes if it is close to an indicator function),

$$\frac{d}{dt}\int_\Omega w_t(x)dx = \frac{d}{dt}\int_\Omega w_0(\varphi^{-1}(x))dx$$

At this point let's make a change of variables $y = \varphi^{-1}(x)$ so $x = \varphi(y)$ and $dx = |D\varphi(y)|dy$

$$= \frac{d}{dt}\int_\Omega w_0(y)|D\varphi(y)|dy = \int_\Omega w_0(y)\frac{d}{dt}|D\varphi(y)|dy$$

On the other hand I could take the time derivative directly

$$= \int_\Omega \frac{d}{dt}w_0(\varphi^{-1}(x))dx = \int_\Omega Dw_0(\varphi^{-1}(x))\frac{d}{dt}\varphi^{-1}(x)dx$$

Using C.2 for the time derivative of the inverse, and then the divergence theorem gives

$$= -\int Dw_0(\varphi^{-1}(x))D\varphi^{-1}vdx$$

$$= -\int D[w_t(\varphi^{-1})]vdx$$

$$= \int w_t(\varphi^{-1})\operatorname{div} vdx$$

Changing variables again gives

$$= \int w_0 \operatorname{div} v(\varphi)|D\varphi|dx$$

So we have

$$\int w_0(y) \frac{d}{dt} |D\varphi(y)| dy = \int w_0(y) \operatorname{div} v(\varphi(y)) |D\varphi(y)| dy$$

Since this is true for all w (weak equality) we have

$$\frac{d}{dt} |D\varphi| = \operatorname{div} v(\varphi) |D\varphi| \tag{C.5}$$

This gives us a nice equation for the log

$$\frac{d}{dt} \log |D\varphi| = \operatorname{div} v(\varphi) \tag{C.6}$$

C.6 Kernel symmetry

Because kernels are used to define inner products, they must be symmetric

$$K(x, y) = K(y, x)$$

This allows us to rearrange expressions involving their derivatives. For a perturbation in direction $h \in \mathbb{R}^3$ we have

$$\begin{aligned} D_1 K(x, y) h &= \lim_{\epsilon \to 0} \frac{d}{d\epsilon} \frac{1}{\epsilon} [K(x + \epsilon h, y) - K(x, y)] \\ &= \lim_{\epsilon \to 0} \frac{d}{d\epsilon} \frac{1}{\epsilon} [K(y, x + \epsilon h) - K(y, x)] \\ &= D_2 K(y, x) h \end{aligned}$$

For this to hold for any perturbation we have the identity

$$D_1 K(x, y) = D_2 K(y, x) \tag{C.7}$$

For translation invariant kernels, $K(x, y) = K(x - y, 0)$, we have

$$\begin{aligned} D_1 K(x, y) h &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} [K(x + \epsilon h, y) - K(x, y)] h \\ &= \lim_{\epsilon \to 0} \frac{d}{d\epsilon} \frac{1}{\epsilon} [K(x + \epsilon h - y, 0) - K(x, y)] h \\ &= \lim_{\epsilon \to 0} \frac{d}{d\epsilon} \frac{1}{\epsilon} [K(x, y - \epsilon h) - K(x, y)] h \\ &= -D_2 K(x, y) h \end{aligned}$$

So for translation invariant kernels we have the identity

$$D_1 K(x, y) = D_2 K(y, x) = -D_2 K(x, y) \tag{C.8}$$

C.7 Completing the square with vectors

For some vector x, square symmetric matrix A, vector a and scalar b, we consider the quantity $x^T Ax + x^T a + b$. We would like to write it in the form $(x - c)^T C(x - c) + d$ for some vector c and scalar d. This will be useful because the exponential of the first term can be easily integrated over (in a Gaussian for example).

Expanding the product gives $x^T Cx - 2x^T Cc + c^T Cc + d$.

Equating quadratic terms gives $A = C$.

Equating linear terms gives $a = -2Cc = -2Ac$ so that $c = -\frac{1}{2}A^{-1}a$.

And last we have $c^T Cc + d = b$ so that $d = b - c^T Cc = b - [-\frac{1}{2}A^{-1}a]^T A[-\frac{1}{2}A^{-1}a] = b - \frac{1}{4}a^T A^{-1}a$

Summarizing, we can write

$$x^T Ax + x^T a + b = (x + \frac{1}{2}A^{-1}a)^T A(x + \frac{1}{2}A^{-1}a) + [b - \frac{1}{4}a^T A^{-1}a]$$

Bibliography

[1] Guy McKhann et al. “Clinical diagnosis of Alzheimer’s disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease”. In: *Neurology* 34.7 (1984), pp. 939–939.

[2] Marshal F Folstein, Susan E Folstein, and Paul R McHugh. “Mini-mental state: a practical method for grading the cognitive state of patients for the clinician”. In: *Journal of psychiatric research* 12.3 (1975), pp. 189–198.

[3] H Braak and E Braak. “Neuropathological stageing of Alzheimer-related changes”. In: *Acta neuropathologica* 82.4 (1991), pp. 239–259.

[4] Reisa A Sperling et al. “Toward defining the preclinical stages of Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease”. In: *Alzheimer’s & dementia* 7.3 (2011), pp. 280–292.

[5] Marilyn S Albert et al. “The diagnosis of mild cognitive impairment due to Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer’s disease”. In: *Alzheimer’s & dementia* 7.3 (2011), pp. 270–279.

[6] Guy M McKhann et al. “The diagnosis of dementia due to Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer’s disease”. In: *Alzheimer’s & dementia* 7.3 (2011), pp. 263–269.

[7] Clifford R Jack et al. “Introduction to the recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease”. In: *Alzheimer’s & Dementia* 7.3 (2011), pp. 257–262.

[8] Ranjan Duara et al. “Insights into cognitive aging and Alzheimers disease using amyloid PET and structural MRI scans”. In: *Clinical and Translational Imaging* 3.1 (2015), pp. 65–74.

[9] John Ashburner and Karl J Friston. “Voxel-based morphometry–the methods”. In: *Neuroimage* 11.6 (2000), pp. 805–821.

[10] Jean Talairach and Pierre Tournoux. “Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging”. In: (1988).

[11] Arno Klein et al. “Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration”. In: *Neuroimage* 46.3 (2009), pp. 786–802.

[12] Roger P Woods et al. “Automated image registration: I. General methods and intrasubject, intramodality validation”. In: *Journal of computer assisted tomography* 22.1 (1998), pp. 139–152.

[13] Roger P Woods et al. “Automated image registration: II. Intersubject validation of linear and nonlinear models”. In: *Journal of computer assisted tomography* 22.1 (1998), pp. 153–165.

[14] Karl Friston et al. “Spatial registration and normalization of images”. In: *Human brain mapping* 3.3 (1995), pp. 165–189.

[15] Jesper LR Andersson, Mark Jenkinson, Stephen Smith, et al. “Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2”. In: *FMRIB Analysis Group of the University of Oxford* 2 (2007).

[16] Bruce Fischl et al. “Sequence-independent segmentation of magnetic resonance images”. In: *Neuroimage* 23 (2004), S69–S84.

[17] J-P Thirion. “Image matching as a diffusion process: an analogy with Maxwell’s demons”. In: *Medical image analysis* 2.3 (1998), pp. 243–260.

[18] Zhark the (Hermite) Cubically Warped. *AFNI 3DQWarp*. `https://afni.nimh.nih.gov/pub/dist/doc/program_help/3dQwarp.html`. 2016.

[19] Michael I Miller, Alain Trouvé, and Laurent Younes. "On the metrics and Euler-Lagrange equations of computational anatomy". In: *Annual review of biomedical engineering* 4.1 (2002), pp. 375–405.

[20] Paul Dupuis, Ulf Grenander, and Michael I Miller. "Variational problems on flows of diffeomorphisms for image matching". In: *Quarterly of applied mathematics* (1998), pp. 587–600.

[21] M Faisal Beg et al. "Computing large deformation metric mappings via geodesic flows of diffeomorphisms". In: *International journal of computer vision* 61.2 (2005), pp. 139–157.

[22] Brian B Avants et al. "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain". In: *Medical image analysis* 12.1 (2008), pp. 26–41.

[23] Michael I Miller, Alain Trouvé, and Laurent Younes. "Geodesic shooting for computational anatomy". In: *Journal of mathematical imaging and vision* 24.2 (2006), pp. 209–228.

[24] Marc Vaillant and Joan Glaunès. "Surface matching via currents". In: *Biennial International Conference on Information Processing in Medical Imaging*. Springer. 2005, pp. 381–392.

[25] Daniel J Tward et al. "Robust diffeomorphic mapping via geodesically controlled active shapes". In: *Journal of Biomedical Imaging* 2013 (2013), p. 3.

[26] Nicolas Charon and Alain Trouvé. "The varifold representation of nonoriented shapes for diffeomorphic registration". In: *SIAM Journal on Imaging Sciences* 6.4 (2013), pp. 2547–2580.

[27] Marc Vaillant et al. "Statistics on diffeomorphisms via tangent space representations". In: *NeuroImage* 23 (2004), S161–S169.

[28] Siu-Wing Cheng, Tamal K Dey, and Jonathan Shewchuk. *Delaunay mesh generation*. CRC Press, 2012.

[29] Keenan Crane et al. “Digital geometry processing with discrete exterior calculus”. In: *ACM SIGGRAPH 2013 Courses*. ACM. 2013, p. 7.

[30] Can Ceritoglu et al. “Computational analysis of LDDMM for brain mapping”. In: *Frontiers in neuroscience* 7 (2013), p. 151.

[31] Timothy F Cootes et al. “Active shape models-their training and application”. In: *Computer vision and image understanding* 61.1 (1995), pp. 38–59.

[32] Anqi Qiu, Laurent Younes, and Michael I Miller. “Principal component based diffeomorphic surface mapping”. In: *IEEE transactions on medical imaging* 31.2 (2012), pp. 302–311.

[33] Bruce Fischl. “FreeSurfer”. In: *Neuroimage* 62.2 (2012), pp. 774–781.

[34] Anqi Qiu and Michael I Miller. “Multi-structure network shape analysis via normal surface momentum maps”. In: *NeuroImage* 42.4 (2008), pp. 1430–1438.

[35] John W Haller et al. “Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas.” In: *Radiology* 202.2 (1997), pp. 504–510.

[36] John G Csernansky et al. “Hippocampal deformities in schizophrenia characterized by high dimensional brain mapping”. In: *American Journal of Psychiatry* 159.12 (2002), pp. 2000–2006.

[37] Joel A Posener et al. “High-dimensional mapping of the hippocampus in depression”. In: *American Journal of Psychiatry* 160.1 (2003), pp. 83–89.

[38] John G Csernansky et al. “Abnormalities of thalamic volume and shape in schizophrenia”. In: *American Journal of Psychiatry* 161.5 (2004), pp. 896–902.

[39] JG Csernansky et al. “Preclinical detection of Alzheimer’s disease: hippocampal shape and volume predict dementia onset in the elderly”. In: *Neuroimage* 25.3 (2005), pp. 783–792.

[40] Michelle M Mielke et al. “Plasma ceramides are altered in mild cognitive impairment and predict cognitive decline and hippocampal volume loss”. In: *Alzheimer’s & Dementia* 6.5 (2010), pp. 378–385.

[41] Sarang C Joshi and Michael I Miller. "Landmark matching via large deformation diffeomorphisms". In: *IEEE Transactions on Image Processing* 9.8 (2000), pp. 1357–1370.

[42] Arthur P Dempster, Nan M Laird, and Donald B Rubin. "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the royal statistical society. Series B (methodological)* (1977), pp. 1–38.

[43] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. "Snakes: Active contour models". In: *International journal of computer vision* 1.4 (1988), pp. 321–331.

[44] L Palm Dzung, Xu Chenyang, and L Prince Jerry. "A survey of current methods in medical image segmentation". In: Technical report.–Johns Hopkins University, Baltimore, 1998. http://www. csem. duke. edu/seminars/segmenation. pdf. 1998.

[45] Stanley Osher and James A Sethian. "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations". In: *Journal of computational physics* 79.1 (1988), pp. 12–49.

[46] Stan Z Li. *Markov random field modeling in computer vision.* Springer Science & Business Media, 2012.

[47] Fred L Bookstein et al. "Principal warps: Thin-plate splines and the decomposition of deformations". In: *IEEE Transactions on pattern analysis and machine intelligence* 11.6 (1989), pp. 567–585.

[48] Daniel Rueckert et al. "Nonrigid registration using free-form deformations: application to breast MR images". In: *IEEE transactions on medical imaging* 18.8 (1999), pp. 712–721.

[49] Bruce Fischl et al. "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain". In: *Neuron* 33.3 (2002), pp. 341–355.

[50] Xiao Han and Bruce Fischl. "Atlas renormalization for improved brain MR image segmentation across scanner platforms". In: *IEEE transactions on medical imaging* 26.4 (2007), pp. 479–486.

[51] Satyanad Kichenassamy et al. “Gradient flows and geometric active contour models”. In: *Computer Vision, 1995. Proceedings., Fifth International Conference on.* IEEE. 1995, pp. 810–815.

[52] Carey E Priebe, Michael I Miller, and J Tilak Ratnanather. “Segmenting magnetic resonance images via hierarchical mixture modelling”. In: *Computational statistics & data analysis* 50.2 (2006), pp. 551–567.

[53] Timothy F Cootes, Gareth J Edwards, Christopher J Taylor, et al. “Active appearance models”. In: *IEEE Transactions on pattern analysis and machine intelligence* 23.6 (2001), pp. 681–685.

[54] Parametrically Deformable. “Boundary finding with parametrically deformable models”. In: *IEEE transactions on pattern analysis and machine intelligence* 14.11 (1992), p. 1061.

[55] Anuj Srivastava et al. “Statistical shape models using elastic-string representations”. In: *Asian Conference on Computer Vision.* Springer. 2006, pp. 612–621.

[56] Daniel Cremers, Timo Kohlberger, and Christoph Schnörr. “Shape statistics in kernel space for variational image segmentation”. In: *Pattern Recognition* 36.9 (2003), pp. 1929–1943.

[57] Laurent Younes, Felipe Arrate, and Michael I Miller. “Evolutions equations in computational anatomy”. In: *NeuroImage* 45.1 (2009), S40–S50.

[58] Felipe Arrate, J Tilak Ratnanather, and Laurent Younes. “Diffeomorphic active contours”. In: *SIAM journal on imaging sciences* 3.2 (2010), pp. 176–198.

[59] Ulf Grenander and Michael I Miller. “Computational anatomy: An emerging discipline”. In: *Quarterly of applied mathematics* (1998), pp. 617–694.

[60] Andrew Staniforth and Jean Côté. “Semi-Lagrangian integration schemes for atmospheric models-a review”. In: *Monthly weather review* 119.9 (1991), pp. 2206–2223.

[61] Daniel Tward et al. “Parametric Surface Diffeomorphometry for Low Dimensional Embeddings of Dense Segmentations and Imagery”. In: *IEEE transactions on pattern analysis and machine intelligence* (2016).

[62] *ADNI*. `http://adni.loni.usc.edu/dataset`. URL: `http://adni.loni.usc.edu/dataset`.

[63] *PREDICT-HD*. `https://www.predict-hd.net/`. URL: `https://www.predict-hd.net/`.

[64] *BIOCARD*. `http://www.alzresearch.org/biocard.cfmdataset`. URL: `http://www.alzresearch.org/biocard.cfmdataset`.

[65] Jun Ma, Michael I Miller, and Laurent Younes. "A bayesian generative model for surface template estimation". In: *Journal of Biomedical Imaging* 2010 (2010), p. 16.

[66] Michael I Miller et al. "The diffeomorphometry of temporal lobe structures in preclinical Alzheimer's disease". In: *NeuroImage: Clinical* 3 (2013), pp. 352–360.

[67] Laurent Younes et al. "Regionally selective atrophy of subcortical structures in prodromal HD as revealed by statistical shape analysis". In: *Human brain mapping* 35.3 (2014), pp. 792–809.

[68] Jorge L Bernal-Rusiel et al. "Statistical analysis of longitudinal neuroimage data with linear mixed effects models". In: *Neuroimage* 66 (2013), pp. 249–260.

[69] Thomas Nichols and Satoru Hayasaka. "Controlling the familywise error rate in functional neuroimaging: a comparative review". In: *Statistical methods in medical research* 12.5 (2003), pp. 419–446.

[70] Vincent Caselles et al. "Minimal surfaces based object segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.4 (1997), pp. 394–398.

[71] Ravi Malladi, James A Sethian, and Baba C Vemuri. "Shape modeling with front propagation: A level set approach". In: *IEEE transactions on pattern analysis and machine intelligence* 17.2 (1995), pp. 158–175.

[72] Anthony Yezzi et al. "A geometric snake model for segmentation of medical imagery". In: *IEEE Transactions on medical imaging* 16.2 (1997), pp. 199–209.

[73] Facundo Mémoli, Guillermo Sapiro, and Paul Thompson. "Implicit brain imaging". In: *NeuroImage* 23 (2004), S179–S188.

[74] Ulf Grenander and Michael I Miller. “Representations of knowledge in complex systems”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1994), pp. 549–603.

[75] Tony F Chan and Luminita A Vese. “Active contours without edges”. In: *IEEE Transactions on image processing* 10.2 (2001), pp. 266–277.

[76] Daniel Cremers et al. “Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional”. In: *International journal of computer vision* 50.3 (2002), pp. 295–313.

[77] Ulf Grenander, Yun-shyong Chow, and Daniel M Keenan. *Hands: A pattern theoretic study of biological shapes.* Vol. 2. Springer Science & Business Media, 2012.

[78] Song Chun Zhu and Alan Yuille. “Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation”. In: *IEEE transactions on pattern analysis and machine intelligence* 18.9 (1996), pp. 884–900.

[79] Oleg Alexandrov and Fadil Santosa. “A topology-preserving level set method for shape optimization”. In: *Journal of Computational Physics* 204.1 (2005), pp. 121–130.

[80] Yunmei Chen et al. “Using prior shapes in geometric active contours in a variational framework”. In: *International Journal of Computer Vision* 50.3 (2002), pp. 315–328.

[81] Oleg Michailovich, Yogesh Rathi, and Allen Tannenbaum. “Image segmentation using active contours driven by the Bhattacharyya gradient flow”. In: *IEEE Transactions on Image Processing* 16.11 (2007), pp. 2787–2801.

[82] Nikos Paragios. “A level set approach for shape-driven segmentation and tracking of the left ventricle”. In: *IEEE transactions on medical imaging* 22.6 (2003), pp. 773–776.

[83] Ganesh Sundaramoorthi and Anthony Yezzi. “Global regularizing flows with topology preservation for active contours and polygons”. In: *IEEE Transactions on Image Processing* 16.3 (2007), pp. 803–812.

[84] Daniel Cremers et al. "A probabilistic level set formulation for interactive organ segmentation". In: *Medical Imaging.* International Society for Optics and Photonics. 2007, pp. 65120V–65120V.

[85] Lixu Gu and Terry Peters. "3D segmentation of medical images using a fast multistage hybrid algorithm". In: *International Journal of Computer Assisted Radiology and Surgery* 1.1 (2006), pp. 23–31.

[86] Carole Le Guyader and Luminita A Vese. "Self-repelling snakes for topology-preserving segmentation models". In: *IEEE Transactions on Image Processing* 17.5 (2008), pp. 767–779.

[87] Mikael Rousson and Daniel Cremers. "Efficient kernel density estimation of shape and intensity priors for level set segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer. 2005, pp. 757–764.

[88] T Schwarz et al. "3D heart segmentation and volumetry using deformable shape models". In: *2007 Computers in Cardiology.* IEEE. 2007, pp. 741–744.

[89] François-Xavier Vialard et al. "Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation". In: *International Journal of Computer Vision* 97.2 (2012), pp. 229–241.

[90] Stanley Durrleman, Stéphanie Allassonnière, and Sarang Joshi. "Sparse adaptive parameterization of variability in image ensembles". In: *International Journal of Computer Vision* 101.1 (2013), pp. 161–183.

[91] Daniel J Tward et al. "Changes in medial temporal lobe anatomy quantified using probabilistic atlas construction and surface diffeomorphometry". In: *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 11.7 (2015), P49–P50.

[92] Jun Ma et al. "Bayesian template estimation in computational anatomy". In: *NeuroImage* 42.1 (2008), pp. 252–261.

[93] John Ashburner and Karl J Friston. "Computing average shaped tissue probability templates". In: *Neuroimage* 45.2 (2009), pp. 333–341.

[94] Arnold Bakker et al. “Response of the medial temporal lobe network in amnestic mild cognitive impairment to therapeutic intervention assessed by fMRI and memory task performance”. In: *NeuroImage: Clinical* 7 (2015), pp. 688–698.

[95] Herman Buschke and Paula Altman Fuld. “Evaluating storage, retention, and retrieval in disordered memory and learning”. In: *Neurology* 24.11 (1974), pp. 1019–1019.

[96] David Wechsler. *WMS-R: Wechsler memory scale-revised: manual.* Psychological Corporation, 1984.

[97] AL Benton. “TheRevised Visual Retention Test”. In: *New York, Psychological Corp* (1963).

[98] Ronald C Petersen. “Mild cognitive impairment as a diagnostic entity”. In: *Journal of internal medicine* 256.3 (2004), pp. 183–194.

[99] H. Duvernoy. *The Human Hippocampus: Functional Anatomy, Vascularization, and Serial Sections with MRI.* Springer Verlag, 2005.

[100] Michael I Miller et al. “Amygdalar atrophy in symptomatic Alzheimer’s disease based on diffeomorphometry: the BIOCARD cohort”. In: *Neurobiology of aging* 36 (2015), S3–S10.

[101] Xiaoying Tang et al. “Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and Alzheimer’s disease: detecting, quantifying, and predicting”. In: *Human brain mapping* 35.8 (2014), pp. 3701–3725.

[102] Laurent Younes et al. “Inferring changepoint times of medial temporal lobe morphometric change in preclinical Alzheimer’s disease”. In: *NeuroImage: Clinical* 5 (2014), pp. 178–187.

[103] Michael I Miller, Laurent Younes, and Alain Trouvé. “Diffeomorphometry and geodesic positioning systems for human anatomy”. In: *Technology* 2.01 (2014), pp. 36–43.

[104] Martin Reuter et al. “Within-subject template estimation for unbiased longitudinal image analysis”. In: *Neuroimage* 61.4 (2012), pp. 1402–1418.

[105] Stanley Durrleman et al. “Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data”. In: *International journal of computer vision* 103.1 (2013), pp. 22–59.

[106] P Thomas Fletcher. "Geodesic regression and the theory of least squares on Riemannian manifolds". In: *International journal of computer vision* 105.2 (2013), pp. 171–185.

[107] Nikhil Singh et al. "Hierarchical geodesic models in diffeomorphisms". In: *International Journal of Computer Vision* 117.1 (2016), pp. 70–92.

[108] Michael I Miller, Alain Trouvé, and Laurent Younes. "Hamiltonian Systems and Optimal Control in Computational Anatomy: 100 Years Since D'Arcy Thompson". In: *Annual review of biomedical engineering* 17 (2015), pp. 447–509.

[109] Daniel J. Tward et al. "Reducing variability in anatomcal definitions over time using longitudinal diffeomorphic mapping". In: *Spectral & Shape Analysis in Medical Imaging, held in conjunction with MICCAI.* Springer, 2016.

[110] D. J. Tward et al. "Local atrophy of entorhinal and trans-entorhinal cortex in mild cognitive impairment measured via diffeomorphometry". In: *Society for Neuroscience 2016 meeting.* 2016. Chap. Session number 040: Diagnostic biomarkers for Alzheimer's disease, Abstract Control Number 8556.

[111] Miaomiao Zhang, William M Wells III, and Polina Golland. "Low-Dimensional Statistics of Anatomical Variability via Compact Representation of Image Deformations". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer. 2016, pp. 166–173.

[112] Andreia V Faria et al. "Linking white matter and deep gray matter alterations in premanifest Huntington disease". In: *NeuroImage: Clinical* 11 (2016), pp. 450–460.

[113] Daniel J Tward et al. "Patient specific dosimetry phantoms using multichannel LDDMM of the whole body". In: *Journal of Biomedical Imaging* 2011 (2011), p. 3.

[114] Saurabh Jain et al. "Computational anatomy gateway: leveraging xsede computational resources for shape analysis". In: *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment.* ACM. 2014, p. 54.

[115] Daniel Tward et al. “OpenCL acceleration of Large Deformation Diffeomorphic Metric Mapping”. In: *The 7th International Workshop on High Performance Computing for Biomedical Image Analysis, held in Conjunction with MICCAI 2014*. 2014.

[116] Daniel Tward et al. “GPU Acceleration on the Stampede Cluster for the Computational Anatomy Gateway”. In: *XSEDE15*. 2015. URL: `https://conferences.xsede.org/-/gpu-acceleration-on-the-stampede-cluster-for-the-computational-anatomy-gateway?redirect=https%3A%2F%2Fconferences.xsede.org%2Fxsede15%2Fschedule%2Fposters%3Fp_p_id%3D101_INSTANCE_zN6kLHVwN7yP%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-2%26p_p_col_count%3D2`.

[117] Daniel J Tward et al. “Tools for studying populations and timeseries of neuroanatomy enabled through GPU acceleration in the Computational Anatomy Gateway”. In: *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale*. ACM. 2016, p. 15.

[118] Kwame S Kutten et al. “Deformably registering and annotating whole CLARITY brains to an atlas via masked LDDMM”. In: *SPIE Photonics Europe*. International Society for Optics and Photonics. 2016, pp. 989616–989616.

[119] Mattias P Heinrich et al. “MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration”. In: *Medical Image Analysis* 16.7 (2012), pp. 1423–1435.

[120] Sureerat Reaungamornrat et al. “MIND demons: symmetric diffeomorphic deformable registration of MR and CT for image-guided spine surgery”. In: *IEEE Trans. Med. Imaging* 35 (2016), pp. 2413–24.

Curriculum Vitae

Daniel Tward was born in Toronto, Ontario, Canada on February 25, 1985. He graduated from the University of Toronto with an Honours Bachelor of Science conferred with high distinction in 2008. His specialization was in biophysics, a program developed jointly by the departments of physics and physiology. Before beginning graduate school, Daniel worked as a research assistant in the Image Guided Therapy Group at the Ontario Cancer Institute, studying physical models of 3D X-ray imaging systems, and psychophysical models of human observer performance under the supervision of Dr. Jeffrey Siewerdsen.

In 2009 Daniel began studying biomedical engineering at John's Hopkins University in Baltimore, Maryland, U.S.A, where he joined the Center for Imaging Science under the supervision of Dr. Michael Miller. Here his focus was in the field of computational anatomy, developing mapping technologies for analysis of neuroimaging data in the presence of noise and other anomalies. His teaching experience includes linear systems, medical imaging, mentoring students through imaging research projects, and developing a new course about computational anatomy, taught for the first time in 2015 and again in 2016. In 2017 Daniel graduated with a doctorate in biomedical engineering.